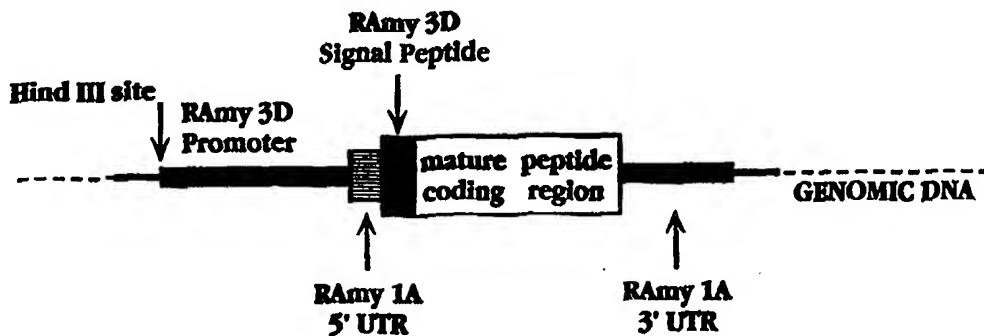




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :	A1	(11) International Publication Number: WO 98/36085
C12N 15/82, 15/57, 15/15, 15/14, C12P 21/02		(43) International Publication Date: 20 August 1998 (20.08.98)
(21) International Application Number: PCT/US98/03068		(72) Inventors; and
(22) International Filing Date: 13 February 1998 (13.02.98)		(75) Inventors/Applicants (for US only): SUTLIFF, Thomas, D. [US/US]; 3405 Cook Street, Rocklin, CA 94765 (US). RODRIGUEZ, Raymond, L. [US/US]; 3017 Anza Boulevard, Davis, CA 95616 (US).
(30) Priority Data:		(74) Agents: PETITHORY, Joanne, R. et al.; Dehlinger & Associates, LLP, P.O. Box 60850, Palo Alto, CA 94306-0850 (US).
60/038,169 13 February 1997 (13.02.97) US		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
60/037,991 13 February 1997 (13.02.97) US		
60/038,170 13 February 1997 (13.02.97) US		
60/038,168 13 February 1997 (13.02.97) US		
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications		
US 60/038,169 (CIP)		
Filed on 13 February 1997 (13.02.97)		
US 60/037,991 (CIP)		
Filed on 13 February 1997 (13.02.97)		
US 60/038,170 (CIP)		
Filed on 13 February 1997 (13.02.97)		
US 60/038,168 (CIP)		
Filed on 13 February 1997 (13.02.97)		
(71) Applicant (for all designated States except US): APPLIED PHYTOLOGICS, INC. [US/US]; 4110 North Freeway Boulevard, Sacramento, CA 95834-1219 (US).		Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PRODUCTION OF MATURE PROTEINS IN PLANTS



(57) Abstract

A method for producing one of the following proteins in transgenic monocot plant cells is disclosed: (i) mature, glycosylated α_1 -antitrypsin (AAT) having the same N-terminal amino acid sequence as mature AAT produced in humans and a glycosylation pattern which increases serum halflife substantially over that of mature non-glycosylated AAT; (ii) mature, glycosylated antithrombin III (ATIII) having the same N-terminal amino acid sequence as mature ATIII produced in humans; (iii) mature human serum albumin (HSA) having the same N-terminal amino acid sequence as mature HSA produced in humans and having the folding pattern of native mature HSA as evidenced by its bilirubin-binding characteristics; and (iv) mature, active subtilisin BPN' (BPN') having the same N-terminal amino acid sequence as BPN' produced in *Bacillus*. Monocot plants cells are transformed with a chimeric gene which includes a DNA coding sequence encoding a fusion protein having an (i) N-terminal moiety corresponding to a rice α -amylase signal sequence peptide and, (ii) immediately adjacent the C-terminal amino acid of said peptide, a protein moiety corresponding to the mature protein to be produced.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

Production of Mature Proteins in Plants

Field of the Invention

The present invention relates to the production of mature proteins in plant cells, and in particular, to the production of proteins in mature secreted form.

5

Background of the Invention

A major commercial focus of biotechnology is the recombinant production of proteins, including both industrial enzymes and proteins that have important therapeutic uses.

Therapeutic proteins are commonly produced recombinantly by microbial expression systems, such as in *E. coli* and the yeast system *S. cerevisiae*. To date, the cost of recombinant proteins produced in a microbial host has limited the availability of a variety of therapeutically important proteins, such as human serum albumin (HSA) and α_1 -antitrypsin (AAT), to the extent that the proteins are in short supply.

Some therapeutic proteins appear to rely on glycosylation for optimal activity or stability, and the general inability of microbial systems to glycosylate or properly glycosylate mammalian proteins has also limited the usefulness of these recombinant expression systems. In some cases, proper protein folding cannot take place, because of the need for mammalian-specific foldases or other folding conditions.

To some extent, protein expression in cultured mammalian cells, or in transgenic animals may overcome the limitations of microbial expression systems. However, the cost per weight ratio of the protein is still high in mammalian expression systems, and the risk of protein contamination by mammalian viruses may be a significant regulatory problem. Protein production by transgenic animals also carries the risk of genetic variation from one generation to another. The attendant risk is variation in the recombinant protein produced, for example, variation in protein processing to yield a nature active protein with different N-terminal residue.

It would therefore be desirable to produce selected therapeutic and industrial proteins in a protein expression system that largely overcomes problems associated with microbial and mammalian-cell systems. In particular, production of the proteins should allow large volume production at low cost, and yield properly processed and glycosylated proteins. The production system should also have a relatively stable genotype from generation to generation. These aims are achieved, in the present invention, for the therapeutic proteins AAT, HSA, and antithrombin III (ATIII), and the industrial enzyme subtilisin BPN'.

Human α_1 -antitrypsin

Human α_1 -antitrypsin (AAT) is a monomer with a molecular weight of about 52Kd. Normal AAT contains 394 residues, with three complex oligosaccharide units exposed to the surface of the

5 of the

molecule, linked to asparagines 46, 83, and 247 (Carrell, P., *et al.*, *Nature* (1982) 298:329).

AAT is the major plasma proteinase inhibitor whose primary function is to control the proteolytic activity of trypsin, elastase, and chymotrypsin in plasma. In particular, the protein is a potent inhibitor of neutrophil elastase, and a deficiency of AAT has been observed in a number of patients with chronic emphysema of the lungs. A proportion of individuals with serum deficiency of AAT may progress to cirrhosis and liver failure (e.g., Wu, Y., *et al.*, *BioEssays* 13(4):163 (1991)).

Because of the key role of AAT as an elastase inhibitor, and because of the prevalence of genetic diseases resulting in deficient serum levels of AAT, there has been an active interest in recombinant synthesis of AAT, for human therapeutic use. To date, this approach has not been satisfactory for AAT produced by recombinant methods, for the reasons discussed above.

Human Antithrombin III

Antithrombin III (ATIII) is the major inhibitor of thrombin and factor Xa, and to a lesser extent, other serine proteases generated during the coagulation process, e.g., factors IXa, XIa, and XIIa. The inhibitory effect of ATIII is accelerated dramatically by heparin. In patients with a history of deep vein thrombosis and pulmonary embolism, the prevalence of ATIII deficiency is 2-3%.

ATIII protein has been useful in treating hereditary ATIII deficiency and has wide clinical applications for the prevention of thrombosis in high risk situations, such as surgery and delivery, and for treating acute thrombotic episodes, when used in combination with heparin.

ATIII is a glycoprotein with a molecular weight of 58,200, having 432 amino acids and containing three disulfide linkages and four asparagine-linked biantennary carbohydrate chains. Because of the key role of ATIII as an anti-thrombotic agent, and because of the broad clinical potential in anti-thrombosis therapy, there has been an active interest in recombinant synthesis of ATIII, for human therapeutic use. To date, this approach has not been satisfactory for ATIII produced by microbial or mammalian recombinant methods, for the reasons discussed above.

Human Serum Albumin

Serum albumin is the main protein component of plasma. Its main function is regulation of colloidal osmotic pressure in the bloodstream. Serum albumin binds numerous ions and small molecules, including Ca^{2+} , Na^+ , K^+ , fatty acids, hormones, bilirubin and certain drugs.

Human serum albumin (HSA) is expressed as a 609 amino acid prepro-protein which is further processed by removal of an amino-terminal peptide and an additional six amino acid residues to form the mature protein. The mature protein found in human serum is a monomeric, unglycosylated protein 585 amino acids in length (66 kDa), with a globular structure maintained by 5 17 disulfide bonds. The pattern of disulfide links forms a structural unit of one small and two large disulfide-linked double loops (Geisow, M.J. *et al.* (1977) *Biochem. J.* 163:477-484) which forms a high-affinity bilirubin binding site.

10 HSA is used to expand blood volume and raise low blood protein levels in cases of shock, trauma, and post-surgical recovery. HSA is often administered in emergency situations to stabilize blood pressure.

15 Because of the key role of HSA as an osmotic stabilizing agent, and because of its broad clinical potential in, *e.g.*, plasma replacement therapy, there has been an active interest in recombinant synthesis of HSA for human therapeutic use. This approach has not been satisfactory for HSA produced by microbial or mammalian recombinant methods, for the reasons discussed above.

Subtilisin BPN'

20 Subtilisin BPN' (BPN') is an important industrial enzyme, particularly for use as a detergent enzyme. Several groups have reported amino acid substitution modifications of the enzyme that are effective in enhancing the activity, pH optimum, stability and/or therapeutic use of the enzyme.

BPN' is expressed in as a 381 amino acid preproenzyme, including 35 amino acid sequence required for secretion and a 77 amino acid moiety which serves as a chaperon to facilitate folding. Studies indicate that the pro moiety acts in trans outside of cells.

25 To date, large-scale production of BPN' is predominantly by microbial fermentation, which has relatively high costs associated with it. In addition, the enzyme tends to auto-degrade at optimal fermentation growth-medium conditions.

Summary of the Invention

30 In one aspect, the invention includes a method of producing, in monocot plant cells, a mature heterologous protein selected from the group consisting of (i) mature, glycosylated α_1 -antitrypsin (AAT) having the same N-terminal amino acid sequence as mature AAT produced in humans and a glycosylation pattern which increases serum halflife substantially over that of non-glycosylated mature AAT; (ii) mature, glycosylated antithrombin III (ATIII) having the same N-terminal amino acid sequence as mature ATIII produced in humans; (iii) mature human serum 35 albumin (HSA) having the same N-terminal amino acid sequence as mature HSA produced in

humans and having the folding pattern of native mature HSA as evidenced by its bilirubin-binding characteristics; and (iv) mature, active subtilisin BPN' (BPN'), glycosylated or non-glycosylated, having the same N-terminal amino acid sequence as BPN' produced in *Bacillus*.

The method includes obtaining monocot cells transformed with a chimeric gene having (i) a 5 monocot transcriptional regulatory region, inducible by addition or removal of a small molecule, or during seed maturation, (ii) a first DNA sequence encoding the heterologous protein, and (iii) a second DNA sequence encoding a signal peptide. The second DNA sequence is operably linked to the transcriptional regulatory region and to the first DNA sequence. The first DNA sequence is in translation-frame with the second DNA sequence, and the two sequences encode a fusion protein. 10 The transformed cells are cultivated under conditions effective to induce the transcriptional regulatory region, thereby promoting expression of the fusion protein and secretion of the mature heterologous protein from the transformed cells. The mature heterologous protein produced by the transformed cells is then isolated.

In one embodiment of the method, the first DNA sequence encodes pro-subtilisin BPN' (proBPN'), the cultivating includes cultivating the transformed cells at a pH between 5 and 6, and the isolating step includes incubating the proBPN' to under condition effective to allow its 15 autoconversion to active mature BPN'. In another embodiment, the first DNA sequence encodes mature BPN', and the cells are transformed with a second chimeric gene containing (i) a transcriptional regulatory region inducible by addition or removal of a small molecule, (ii) a third DNA sequence encoding the pro-peptide moiety of BPN', and (iii) a fourth DNA sequence encoding a 20 signal polypeptide. The fourth DNA sequence is operably linked to the transcriptional regulatory region and to the third DNA sequence, and the signal polypeptide is in translation-frame with the pro-peptide moiety and is effective to facilitate secretion of expressed pro-peptide moiety from the transformed cells. The cultivating step includes cultivating the transformed cells at a pH between 5 25 and 6, and the isolating step includes incubating the mature BPN' and the pro-moiety under conditions effective to allow the conversion of BPN' by the pro- moiety to active mature BPN'.

In another embodiment of the method, the signal peptide is the RAmy3D signal peptide (SEQ ID NO:1) or the RAmy1A signal peptide (SEQ ID NO:4). The coding sequence of the signal peptide may be a codon-optimized sequence, such as the codon-optimized RAmy3D sequence 30 identified as SEQ ID NO:3. The first DNA sequence may also be codon-optimized. Exemplary codon-optimized signal peptide-heterologous protein fusion protein coding sequences include 3D-AAT (SEQ ID NO:18), 3D-ATIII (SEQ ID NO:19), and 3D-HSA (SEQ ID NO:20). The first DNA sequence may further contain codon substitutions which eliminate one or more potential glycosylation sites present in the native amino acid sequence of the heterologous protein, such as the 35 codon-optimized sequence encoding 3D-proBPN' (SEQ ID NO:21).

In other embodiments of the method, the transcriptional regulatory region may be a promoter derived from a rice or barley α -amylase gene, including RAmy1A, RAmy1B, RAmy2A, RAmy3A, RAmy3B, RAmy3C, RAmy3D, RAmy3E, pM/C, gKAmy141, gKAmy155, Amy32b, or HV18. The chimeric gene may further include, between the transcriptional regulatory region and the fusion protein coding sequence, the 5' untranslated region (5' UTR) of an inducible monocot gene such as one of the rice or barley α -amylase genes described above. One preferred 5' UTR is that from the RAmy1A gene, which is effective to enhance the stability of the gene transcript. The chimeric gene may further include, downstream of the coding sequence, the 3' untranslated region (3' UTR) from an inducible monocot gene, such as one of the rice or barley α -amylase genes mentioned above. One preferred 3' UTR is from the RAmy1A gene.

Where the method is employed in protein production in a monocot cell culture, preferred promoters are the RAmy3D and RAmy3E gene promoters, which are upregulated by sugar depletion in cell culture. Where the gene is employed in protein production in germinating seeds, a preferred promoter is the RAmy1A gene promoter, which is upregulated by gibberellic acid during seed germination. Where gene is upregulated during seed maturation, a preferred promoter is the barley endosperm-specific B1-hordein promoter.

The invention also includes a mature heterologous protein produced by the above method. The protein has a glycosylation pattern characteristic of the monocot plant in which the protein is produced. The glycosylated protein is selected from the group consisting of (i) mature glycosylated α_1 -antitrypsin (AAT) having the same N-terminal amino acid sequence as mature AAT produced in humans and having a glycosylation pattern which increases serum halflife substantially over that of non-glycosylated mature AAT; (ii) mature glycosylated antithrombin III (ATIII) having the same N-terminal amino acid sequence as mature ATIII produced in humans; and (iii) mature glycosylated subtilisin BPN' (BPN') having the same N-terminal amino acid sequence as BPN' produced in *Bacillus*.

The invention also includes plant cells and seeds capable of producing the mature heterologous proteins according to the above method.

These and other objects and features of the invention will be more fully understood when the following detailed description of the invention is read in conjunction with the accompanying drawings.

Brief Description of the Figures

Fig. 1 shows, in the lower row, the amino acid sequence of a RAmy3D signal sequence portion employed in the invention, identified as SEQ ID NO:1; in the middle row, the corresponding native coding sequence, identified as SEQ ID NO:2; and in the upper row, a corresponding codon-optimized sequence, identified as SEQ ID NO:3;

Fig. 2 illustrates the components of a chimeric gene constructed in accordance with an embodiment of the invention;

5 Figs. 3A and 3B illustrate the construction of an exemplary transformation vector for use in transforming a monocot plant, for production of a mature protein in cell culture in accordance with one embodiment of the invention (native mature AAT coding sequence under control of the RAmy3D promoter and signal sequence);

Fig. 4 illustrates factors in the metabolic regulation of AAT production in rice cell culture;

Fig. 5 shows immunodetection of AAT using antibody raised against the C-terminal region of AAT;

10 Fig. 6 shows Western blot analysis of AAT produced by transformed rice cell lines 18F, 11B, and 27F;

Fig. 7 shows the time course of elastase:AAT complex formation in human and rice-produced forms of AAT;

15 Fig. 8 shows an N-terminal sequence for mature α_1 -antitrypsin (AAT) produced in accordance with the invention, identified herein as SEQ ID NO:22;

Fig. 9 shows a Western blot of ATIII produced in accordance with the invention;

Fig. 10 shows a Western blot of plant-produced BPN', comparing expression from codon-optimized and native coding sequences;

20 Fig. 11 compares the specific activity of BPN' codon-optimized (AP106) vs. BPN' native (AP101) expression in rice callus cell culture; and

Fig. 12 shows a western blot of HSA produced in germinating seeds in accordance with the invention.

Brief Description of the Sequences

25 SEQ ID NO:1 is the amino acid sequence of the RAmy3D signal peptide;

SEQ ID NO:2 is the native sequence encoding the RAmy3D signal peptide;

SEQ ID NO:3 is a codon-optimized sequence encoding the RAmy3D signal peptide;

SEQ ID NO:4 is the amino acid sequence of the RAmy1A signal peptide;

SEQ ID NO:5 is the 5' UTR derived from the RAmy1A gene;

30 SEQ ID NO:6 is the 3' UTR derived from the RAmy1A gene;

SEQ ID NO:7 is the amino acid sequence of mature α_1 -antitrypsin (AAT);

SEQ ID NO:8 is the native DNA coding sequence of mature AAT;

SEQ ID NO:9 is the amino acid sequence of mature antithrombin III (ATIII);

SEQ ID NO:10 is the native DNA coding sequence of mature ATIII;

35 SEQ ID NO:11 is the amino acid sequence of mature human serum albumin (HSA);

SEQ ID NO:12 is the native DNA coding sequence of mature HSA;

SEQ ID NO:13 is the amino acid sequence of native proBPN';

SEQ ID NO:14 is the native DNA coding sequence of proBPN';

SEQ ID NO:15 is the amino acid sequence of the "pro" moiety of BPN';

5 SEQ ID NO:16 is the amino acid sequence of native mature BPN';

SEQ ID NO:17 is the amino acid sequence of a mature BPN' variant in which all potential N-glycosylation sites are removed according to Table 2;

SEQ ID NO:18 is a codon-optimized sequence encoding the RAmy3D signal sequence/mature α_1 -antitrypsin fusion protein;

10 SEQ ID NO:19 is a sequence encoding the RAmy3D signal sequence/mature antithrombin III fusion protein, with a codon-optimized RAmy3D coding sequence fused to the native mature ATIII coding sequence;

15 SEQ ID NO:20 is a sequence encoding the RAmy3D signal sequence/mature human serum albumin fusion protein, with a codon-optimized RAmy3D coding sequence fused to the native mature HSA coding sequence;

SEQ ID NO:21 is a codon-optimized sequence encoding the RAmy3D signal sequence/prosubtilisin BPN' fusion protein;

SEQ ID NO:22 is the N-terminal sequence of mature α_1 -antitrypsin produced in accordance with the invention;

20 SEQ ID NO:23 is an oligonucleotide used to prepare the intermediate p3DProSig construct of Example 1;

SEQ ID NO:24 is the complement of SEQ ID NO:23;

SEQ ID NO:25 is an oligonucleotide used to prepare the intermediate p3DProSigENDlink construct of Example 1;

25 SEQ ID NO:26 is the complement of SEQ ID NO:25;

SEQ ID NO:27 is one of six oligonucleotides used to prepare the intermediate p1AProSig construct of Example 1;

SEQ ID NO:28 is one of six oligonucleotides used to prepare the intermediate p1AProSig construct of Example 1;

30 SEQ ID NO:29 is one of six oligonucleotides used to prepare the intermediate p1AProSig construct of Example 1;

SEQ ID NO:30 is one of six oligonucleotides used to prepare the intermediate p1AProSig construct of Example 1;

35 SEQ ID NO:31 is one of six oligonucleotides used to prepare the intermediate p1AProSig construct of Example 1;

SEQ ID NO:32 is one of six oligonucleotides used to prepare the intermediate p1AProSig construct of Example 1;

SEQ ID NO:33 is the N-terminal primer used to PCR-amplify the AAT coding sequence according to Example 1; and

5 SEQ ID NO:34 is the C-terminal primer used to PCR-amplify the AAT coding sequence according to Example 1.

Detailed Description of the Invention

I. Definitions:

10 The terms below have the following meaning, unless indicated otherwise in the specification.

"Cell culture" refers to cells and cell clusters, typically callus cells, growing on or suspended in a suitable growth medium.

15 "Germination" refers to the breaking of dormancy in a seed and the resumption of metabolic activity in the seed, including the production of enzymes effective to break down starches in the seed endosperm.

"Inducible" means a promoter that is upregulated by the presence or absence of a small molecules. It includes both indirect and direct inducement.

20 "Inducible during germination" refers to promoters which are substantially silent but not totally silent prior to germination but are turned on substantially (greater than 25%) during germination and development in the seed. Examples of promoters that are inducible during germination are presented below.

25 "Small molecules", in the context of promoter induction, are typically small organic or bioorganic molecules less than about 1 kDa. Examples of such small molecules include sugars, sugar-derivatives (including phosphate derivatives), and plant hormones (such as, gibberellic or abscisic acid).

30 "Specifically regulatable" refers to the ability of a small molecule to preferentially affect transcription from one promoter or group of promoters (e.g., the α -amylase gene family), as opposed to non-specific effects, such as, enhancement or reduction of global transcription within a cell by a small molecule.

35 "Seed maturation" or "grain development" refers to the period starting with fertilization in which metabolizable reserves, e.g., sugars, oligosaccharides, starch, phenolics, amino acids, and proteins, are deposited, with and without vacuole targeting, to various tissues in the seed (grain), e.g., endosperm, testa, aleurone layer, and scutellar epithelium, leading to grain enlargement, grain filling, and ending with grain desiccation.

"Inducible during seed maturation" refers to promoters which are turned on substantially (greater than 25%) during seed maturation.

5 "Heterologous DNA" or "foreign DNA" refers to DNA which has been introduced into plant cells from another source, or which is from a plant source, including the same plant source, but which is under the control of a promoter or terminator that does not normally regulate expression of the heterologous DNA.

10 "Heterologous protein" is a protein, including a polypeptide, encoded by a heterologous DNA. A "transcription regulatory region" or "promoter" refers to nucleic acid sequences that influence and/or promote initiation of transcription. Promoters are typically considered to include regulatory regions, such as enhancer or inducer elements.

15 A "chimeric gene," in the context of the present invention, typically comprises a promoter sequence operably linked to DNA sequence that encodes a heterologous gene product, *e.g.*, a selectable marker gene or a fusion protein gene. A chimeric gene may also contain further transcription regulatory elements, such as transcription termination signals, as well as translation regulatory signals, such as, termination codons.

"Operably linked" refers to components of a chimeric gene or an expression cassette that function as a unit to express a heterologous protein. For example, a promoter operably linked to a heterologous DNA, which encodes a protein, promotes the production of functional mRNA corresponding to the heterologous DNA.

20 A "product" encoded by a DNA molecule includes, for example, RNA molecules and polypeptides.

"Removal" in the context of a metabolite includes both physical removal as by washing and the depletion of the metabolite through the absorption and metabolizing of the metabolite by the cells.

25 "Substantially isolated" is used in several contexts and typically refers to the at least partial purification of a protein or polypeptide away from unrelated or contaminating components. Methods and procedures for the isolation or purification of proteins or polypeptides are known in the art.

30 "Stably transformed" as used herein refers to a cereal cell or plant that has foreign nucleic acid stably integrated into its genome which is transmitted through multiple generations.

" α_1 -antitrypsin" or "AAT" refers to the protease inhibitor which has an amino acid sequence substantially identical or homologous to AAT protein identified by SEQ ID NO:7.

35 "Antithrombin III" or "ATIII" refers to the heparin-activated inhibitor of thrombin and factor Xa, and which has an amino acid sequence substantially identical or homologous to ATIII protein identified by SEQ ID NO:9.

"Human serum albumin" or "HSA" refers to a protein which has an amino acid sequence substantially identical or homologous to the mature HSA protein identified by SEQ ID NO:11.

5 "Subtilisin" or "subtilisin BPN'" or "BPN'" refers to the protease enzyme produced naturally by *B. amyloliquefaciens*, and having the sequence of SEQ ID NO:16, or a sequence homologous therewith.

"proBPN'" refers to a form of BPN' having an approximately 78 amino-acid "pro" moiety that functions as a chaperon polypeptide to assist in folding and activation of the BPN', and having the sequence in SEQ ID NO:13, or a sequence homologous therewith.

10 "Codon optimization" refers to changes in the coding sequence of a gene to replace native codons with those corresponding to optimal codons in the host plant.

15 A DNA sequence is "derived from" a gene, such as a rice or barley α -amylase gene, if it corresponds in sequence to a segment or region of that gene. Segments of genes which may be derived from a gene include the promoter region, the 5' untranslated region, and the 3' untranslated region of the gene.

15

II. Transformed plant cells

20 The plants used in the process of the present invention are derived from monocots, particularly the members of the taxonomic family known as the Gramineae. This family includes all members of the grass family of which the edible varieties are known as cereals. The cereals include a wide variety of species such as wheat (*Triticum* sps.), rice (*Oryza* sps.) barley (*Hordeum* sps.) oats, (*Avena* sps.) rye (*Secale* sps.), corn (*Zea* sps.) and millet (*Pennisetum* sps.). In the present invention, preferred family members are rice and barley.

25 Plant cells or tissues derived from the members of the family are transformed with expression constructs (i.e., plasmid DNA into which the gene of interest has been inserted) using a variety of standard techniques (e.g., electroporation, protoplast fusion or microparticle bombardment). The expression construct includes a transcription regulatory region (promoter) whose transcription is specifically upregulated by the presence of absence of a small molecule, such as the reduction or depletion of sugar, e.g., sucrose, in culture medium, or in plant tissues, e.g., germinating seeds. In the present invention, particle bombardment is the preferred transformation 30 procedure.

35 The construct also includes a gene encoding a mature heterologous protein in a form suitable for secretion from plant cells. The gene encoding the recombinant heterologous protein is placed under the control of a metabolically regulated promoter. Metabolically regulated promoters are those in which mRNA synthesis or transcription, is repressed or upregulated by a small metabolite or hormone molecule, such as the rice RAmy3D and RAmy3E promoters, which are

upregulated by sugar-depletion in cell culture. For protein production in germinating seeds from regenerated transgenic plants, a preferred promoter is the Ramy 1A promoter, which is up-regulated by gibberellic acid during seed germination. The expression construct also utilizes additional regulatory DNA sequences *e.g.*, preferred codons, termination sequences, to promote efficient 5 translation of AAT, as will be described.

A. Plant Expression Vector

Expression vectors for use in the present invention comprise a chimeric gene (or expression cassette), designed for operation in plants, with companion sequences upstream and downstream 10 from the expression cassette. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from bacteria to the desired plant host. Suitable transformation vectors are described in related application PCT WO 95/14099, published May 25, 1995, which is incorporated by reference herein. Suitable 15 components of the expression vector, including an inducible promoter, coding sequence for a signal peptide, coding sequence for a mature heterologous protein, and suitable termination sequences are discussed below. One exemplary vector is the p3D(AAT)v1.0 vector illustrated in Figs 3A and 3B.

A1. Promoters

The transcription regulatory or promoter region is chosen to be regulated in a manner 20 allowing for induction under selected cultivation conditions, *e.g.*, sugar depletion in culture or water uptake followed by gibberellic acid production in germinating seeds. Suitable promoters, and their method of selection are detailed in above-cited PCT application WO 95/14099. Examples of 25 such promoters include those that transcribe the cereal α -amylase genes and sucrose synthase genes, and are repressed or induced by small molecules, like sugars, sugar depletion or phytohormones such as gibberellic acid or abscisic acid. Representative promoters include the promoters from the 30 rice α -amylase RAmy1A, RAmy1B, RAmy2A, RAmy3A, RAmy3B, RAmy3C, RAmy3D, and RAmy3E genes, and from the pM/C, gKAmy141, gKAmy155, Amy32b, and HV18 barley α -amylase genes. These promoters are described, for example, in ADVANCES IN PLANT BIOTECHNOLOGY Ryu, D.D.Y., *et al*, Eds., Elsevier, Amsterdam, 1994, p.37, and references cited 35 therein. Other suitable promoters include the sucrose synthase and sucrose-6-phosphate-synthetase (SPS) promoters from rice and barley.

Other suitable promoters include promoters which are regulated in a manner allowing for induction under seed-maturation conditions. Examples of such promoters include those associated 35 with the following monocot storage proteins: rice glutelins, oryzins, and prolamines, barley hordeins, wheat gliadins and glutelins, maize zeins and glutelins, oat glutelins, and sorghum

kafirins, millet pennisetins, and rye secalins.

A preferred promoter for expression in germinating seeds is the rice α -amylase RAmy1A promoter, which is upregulated by gibberellic acid. Preferred promoters for expression in cell culture are the rice α -amylase RAmy3D and RAmy3E promoters which are strongly upregulated by sugar depletion in the culture. These promoters are also active during seed germination. A preferred promoter for expression in maturing seeds is the barley endosperm-specific B1-hordein promoter (Brandt, A., *et al.*, (1985) Carlsberg Res. Commun. 50:333-345).

The chimeric gene may further include, between the promoter and coding sequences, the 5' untranslated region (5' UTR) of an inducible monocot gene, such as the 5' UTR derived from one of the rice or barley α -amylase genes mentioned above. One preferred 5' UTR is that derived from the RAmy1A gene, which is effective to enhance the stability of the gene transcript. This 5' UTR has the sequence given by SEQ ID NO:5 herein.

A2. Signal Sequences

In addition to encoding the protein of interest, the chimeric gene encodes a signal sequence (or signal peptide) that allows processing and translocation of the protein, as appropriate. Suitable signal sequences are described in above-referenced PCT application WO 95/14099. One preferred signal sequence is identified as SEQ ID NO:1 and is derived from the RAmy3D promoter. Another preferred signal sequence is identified as SEQ ID NO:4 and is derived from the RAmy1A promoter. The plant signal sequence is placed in frame with a heterologous nucleic acid encoding a mature protein, forming a construct which encodes a fusion protein having an N-terminal region corresponding to the signal peptide and, immediately adjacent to the C-terminal amino acid of the signal peptide, the N-terminal amino acid of the mature heterologous protein. The expressed fusion protein is subsequently secreted and processed by signal peptidase cleavage precisely at the junction of the signal peptide and the mature protein, to yield the mature heterologous protein.

In another embodiment of the invention, the coding sequence in the fusion protein gene, in at least the coding region for the signal sequence, may be codon-optimized for optimal expression in plant cells, *e.g.*, rice cells, as described below. The upper row in Fig. 1 shows one codon-optimized coding sequence for the RAmy3D signal sequence, identified herein as SEQ ID NO:3.

30

A3. Naturally-Occurring Heterologous Protein Coding Sequences

(i) α_1 -Antitrypsin: Mature human AAT is composed of 394 amino acids, having the sequence identified herein as SEQ ID NO:7. The protein has N-glycosylation sites at asparagines 46, 83 and 247. The corresponding native DNA coding sequence is identified herein as SEQ ID NO:8.

(ii) Antithrombin III: Mature human ATIII is composed of 432 amino acids, having the sequence identified herein as SEQ ID NO:9. The protein has N-glycosylation sites at the four asparagine residues 96, 135, 155, and 192. The corresponding native DNA coding sequence is identified herein as SEQ ID NO:10.

5 (iii) Human serum albumin: Mature HSA as found in human serum is composed of 585 amino acids, having the sequence identified herein as SEQ ID NO:11. The protein has no N-linked glycosylation sites. The corresponding native DNA coding sequence is identified herein as SEQ ID NO:12.

10 (iv) Subtilisin BPN': Native proBPN' as produced in *B. amyloliquefaciens* is composed of 352 amino acids, having the sequence identified herein as SEQ ID NO:13. The corresponding native DNA coding sequence is identified herein as SEQ ID NO:14. The proBPN' polypeptide contains a 77 amino acid "pro" moiety which is identified herein as SEQ ID NO:15. The remainder of the polypeptide, which forms the mature active BPN', is a 275 amino acid sequence identified herein by SEQ ID NO:16. Native BPN' as produced in *Bacillus* is not glycosylated.

15

A4. Codon-Optimized Coding Sequences

20 In accordance with one aspect of the invention, it has been discovered that a severalfold enhancement of expression level can be achieved in plant cell culture by modifying the native coding sequence of a heterologous gene by contain predominantly or exclusively, highest-frequency codons found in the plant cell host.

25 The method will be illustrated for expression of a heterologous gene in rice plant cells, it being recognized that the method is generally applicable to any monocot. As a first step, a representative set of known coding gene sequence from rice is assembled. The sequences are then analyzed for codon frequency for each amino acid, and the most frequent codon is selected for each amino acid. This approach differs from earlier reported codon matching methods, in which more than one frequent codon is selected for at least some of the amino acids. The optimal codons selected in this manner for rice and barley are shown in Table 1.

Table 1

Amino Acid	Rice Preferred Codon	Barley Preferred Codon
Ala A	GCC	
Arg R	CGC	
Asn N	AAC	

30

Amino Acid	Rice Preferred Codon	Barley Preferred Codon
Asp D	GAC	
Cys C	UGC	
Gln Q	CAG	
Glu E	GAG	
Gly G	GGC	
His H	CAC	
Ile I	AUC	
Leu L	CUC	
Lys K	AAG	
Phe F	UUC	
Pro P	CCG	CCC
Ser S	AGC	UCC
Thr T	ACC	
Tyr Y	UAC	
Val V	GUC	GUG
stop	UAA	UGA

As indicated above, the fusion protein coding sequence in the chimeric gene is constructed such that the final (C-terminal) codon in the signal sequence is immediately followed by the codon 5 for the N-terminal amino acid in the mature form of the heterologous protein. Exemplary fusion protein genes, in accordance with the present invention, are identified herein as follows:

SEQ ID NO:18, corresponding to codon-optimized coding sequences of the fusion protein consisting of RAmy3D signal sequence/mature α_1 -antitrypsin;

SEQ ID NO:19, corresponding to the fusion protein coding sequence consisting of the 10 codon-optimized RAmy3D signal sequence and the native mature antithrombin III sequence;

SEQ ID NO:20, corresponding to the fusion protein coding sequence consisting of the codon-optimized RAmy3D signal sequence and the native mature human serum albumin sequence;

SEQ ID NO:21, corresponding to codon-optimized coding sequence of the fusion protein RAmy3D signal sequence/prosubtilisin BPN'. In this instance, prosubtilisin is considered the 15 "mature" protein, in that secreted prosubtilisin can autocatalyze to active, mature subtilisin.

In a preferred embodiment, the BPN' coding sequence is further modified to eliminate

potential N-glycosylation sites, as native BPN' is not glycosylated. Table 2 illustrates preferred codon substitutions, which eliminate all potential N-glycosylation sites in subtilisin BPN'. SEQ ID NO:17 corresponds to a mature BPN' amino acid sequence containing the substitutions presented in Table 2.

5

Table 2

N-Glycosylation Sites	Location (Asn) (in mature protein)	Amino Acid Substitution
Asn Asn Ser	61	Thr Asn Ser
Asn Asn Ser	76	Thr Asn Ser
Asn Met Ser	123	Thr Met Ser
Asn Gly Thr	218	Ser Gly Thr ¹
Asn Trp Thr	240	Thr Trp Thr

¹improved thermostability; Bryan, *et al.*, *Proteins: Structure, Function, and Genetics* 1:326 (1986).

10

A5. Transcription and Translation Terminators

The chimeric gene may also include, downstream of the coding sequence, the 3' untranslated region (3' UTR) from an inducible monocot gene, such as one of the rice or barley α -amylase genes mentioned above. One preferred 3' UTR is that derived from the RAmy1A gene, whose sequence is given by SEQ ID NO:6. This sequence includes non-coding sequence 5' to the polyadenylation site, the polyadenylation site, and the transcription termination sequence. The transcriptional termination region may be selected, particularly for stability of the mRNA to enhance expression. Polyadenylation tails (Alber and Kawasaki, 1982, *Mol. and Appl. Genet.* 1:419-434) are also commonly added to the expression cassette to optimize high levels of transcription and proper transcription termination, respectively. Polyadenylation sequences include but are not limited to the *Agrobacterium* octopine synthetase signal (Gielen, *et al.*, *EMBO J.* 3:835-846 (1984) or the nopaline synthase of the same species (Depicker, *et al.*, *Mol. Appl. Genet.* 1:561-573 (1982).

Since the ultimate expression of the heterologous protein will be in a eukaryotic cell (in this case, a member of the grass family), it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicing machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code (Reed and Maniatis, *Cell* 41:95-105 (1985).

Fig. 2 shows the elements of one preferred chimeric gene constructed in accordance with the invention, and intended particularly for use in protein expression in a rice cell suspension culture. The gene includes, in a 5' to 3' direction, the promoter from the RAmy3D gene, which is inducible in cell culture with sugar depletion, the 5' UTR from the RAmy1A gene, which confers enhanced stability on the gene transcript, the RAmy3D signal sequence coding region, as identified above, the coding region of a heterologous protein to be produced, and a 3' UTR region from the RAmy1A gene.

III. Plant Transformation

For transformation of plants, the chimeric gene is placed in a suitable expression vector designed for operation in plants. The vector includes suitable elements of plasmid or viral origin that provide necessary characteristics to the vector to permit the vectors to move DNA from bacteria to the desired plant host. Suitable transformation vectors are described in related application PCT WO 95/14099, published May 25, 1995, which is incorporated by reference herein. Suitable components of the expression vector, including the chimeric gene described above, are discussed below. One exemplary vector is the p3Dv1.0 vector described in Example 1.

A. Transformation Vector

Vectors containing a chimeric gene of the present invention may also include selectable markers for use in plant cells (such as the *nptII* kanamycin resistance gene, for selection in kanamycin-containing or the phosphinothricin acetyltransferase gene, for selection in medium containing phosphinothricin (PPT)).

The vectors may also include sequences that allow their selection and propagation in a secondary host, such as sequences containing an origin of replication and a selectable marker such as antibiotic or herbicide resistance genes, *e.g.*, HPH (Hagio *et al.*, *Plant Cell Reports* **14**:329 (1995); van der Elzer, *Plant Mol. Biol.* **5**:299-302 (1985). Typical secondary hosts include bacteria and yeast. In one embodiment, the secondary host is *Escherichia coli*, the origin of replication is a colE1-type, and the selectable marker is a gene encoding ampicillin resistance. Such sequences are well known in the art and are commercially available as well (*e.g.*, Clontech, Palo Alto, CA; Stratagene, La Jolla, CA).

The vectors of the present invention may also be modified to intermediate plant transformation plasmids that contain a region of homology to an *Agrobacterium tumefaciens* vector, a T-DNA border region from *Agrobacterium tumefaciens*, and chimeric genes or expression cassettes (described above). Further, the vectors of the invention may comprise a disarmed plant tumor inducing plasmid of *Agrobacterium tumefaciens*.

The vector described in Example 1, and having a promoter from the RAmy3D gene, is suitable for use in a method of mature protein production in cell culture, where the RAmy3D promoter is induced by sugar depletion in cell culture medium. Other promoters may be selected for other applications, as indicated above. For example, for mature protein expression in 5 germinating seeds, the coding sequence may be placed under the control of the rice α -amylase RAmy1A promoter, which is inducible by gibberellic acid during seed germination.

B. Transformation of plant cells

Various methods for direct or vectored transformation of plant cells, *e.g.*, plant protoplast 10 cells, have been described, *e.g.*, in above-cited PCT application WO 95/14099. As noted in that reference, promoters directing expression of selectable markers used for plant transformation (*e.g.*, nptII) should operate effectively in plant hosts. One such promoter is the nos promoter from native Ti plasmids (Herrera-Estrella, *et al.*, *Nature* 303:209-213 (1983). Others include the 35S and 19S 15 promoters of cauliflower mosaic virus (Odell, *et al.*, *Nature* 313:810-812 (1985) and the 2' promoter (Velten, *et al.*, *EMBO J.* 3:2723-2730 (1984).

In one preferred embodiment, the embryo and endosperm of mature seeds are removed to exposed scutulum tissue cells. The cells may be transformed by DNA bombardment or injection, or by vectored transformation, *e.g.*, by *Agrobacterium* infection after bombarding the scutellar cells 20 with microparticles to make them susceptible to *Agrobacterium* infection (Bidney *et al.*, *Plant Mol. Biol.* 18:301-313, 1992).

One preferred transformation follows the methods detailed generally in Sivamani, E. *et al.*, *Plant Cell Reports* 15:465 (1996); Zhang, S., *et al.*, *Plant Cell Reports* 15:465 (1996); and Li, L., *et al.*, *Plant Cell Reports* 12:250 (1993). Briefly, rice seeds are sterilized by standard methods, and 25 callus induction from the seeds is carried out on MB media with 2,4D. During a first incubation period, callus tissue forms around the embryo of the seed. By the end of the incubation period, (*e.g.*, 14 days at 28°C) the calli are about 0.25 to 0.5 cm in diameter. Callus mass is then detached from the seed, and placed on fresh NB media, and incubated again for about 14 days at 28°C. After 30 the second incubation period, satellite calli developed around the original "mother" callus mass. These satellite calli were slightly smaller, more compact and defined than the original tissue. It was these calli were transferred to fresh media. The "mother" calli was not transferred. The goal was to select only the strongest, most vigorous growing tissue for further culture.

Calli to be bombarded are selected from 14-day-old subcultures. The size, shape, color and density are all important in selecting calli in the optimal physiological condition for transformation. The calli should be between .8 and 1.1 mm in diameter. The calli should appear as spherical 35 masses with a rough exterior.

Transformation is by particle bombardment, as detailed in the references cited above. After the transformation steps, the cells are typically grown under conditions that permit expression of the selectable marker gene. In a preferred embodiment, the selectable marker gene is HPH. It is preferred to culture the transformed cells under multiple rounds of selection to produce a uniformly 5 stable transformed cell line.

IV. Cell Culture Production of Mature Heterologous Protein

Transgenic cells, typically callus cells, are cultured under conditions that favor plant cell growth, until the cells reach a desired cell density, then under conditions that favor expression of 10 the mature protein under the control of the given promoter. Preferred culture conditions are described below and in Example 2. Purification of the mature protein secreted into the medium is by standard techniques known by those of skill in the art.

Production of mature AAT: In a preferred embodiment, the culture medium contains a phosphate buffer, *e.g.*, the 20 mM phosphate buffer, pH 6.8 described in Example 2, to reduce 15 AAT degradation catalyzed by metals. Alternatively, or in addition, a metal chelating agent, such as EDTA, may be added to the medium.

Following the cell culture method described in Example 2, cell culture media was partially purified and the fraction containing AAT was analyzed by Western blot, as shown in Fig. 4. The first two lanes ("phosphate") show AAT bands both in the presence and absence of elastase ("+E" 20 and "-E"), where the higher molecular weight bands in the presence of elastase correspond roughly to a 58-59 kdal AAT/elastase complex. Also as seen in the figure, expression was high in the absence of sucrose, but nearly undetectable in the presence of sucrose.

To ascertain the degree of glycosylation (as determined by apparent molecular weight by SDS-PAGE) the protein produced in culture was fractionated by SDS-PAGE and immunodetected 25 with a labeled antibody raised against the C-terminal portion of AAT, as shown in Fig. 5. Lane 4 contains human AAT, and its migration position corresponds to about 52 kdal. In lane 3 is the plant-produced AAT, having an apparent molecular weight of about 49-50 kdal, indicating an extent of glycosylation of up to 60-80% of the glycosylation found in human AAT (non-glycosylated AAT has a molecular weight of 45 kdal).

Similar results are shown in the Western blots in Fig. 6. Lanes 1-3 in this figure 30 correspond to decreasing amount (15, 10, and 5 ng) of human AAT; lane 4, to 10 μ l supernatant from a non-expressing plant cell line; lanes 5 and 6, to 10 μ l supernatant from AAT-expressing plant cell lines 11B and 27F, respectively, and lane 7, to 10 μ l supernatant from cell line 27F plus 250 ng trypsin. The upward mobility shift in lane 7 is indicative of association between trypsin and 35 the plant-produced AAT.

The ability of plant-produced AAT to bind to elastase is demonstrated in Fig. 7, which shows the shift in molecular weight over a 30 minute binding interval for the 52 kdal human AAT (lanes 1-4) and the 49-50 kdal plant-produced AAT.

5 To demonstrate that the mature protein is produced in secreted form, with the desired N-terminus, a chimeric gene constructed as above, and having the coding sequence for mature α_1 -antitrypsin was expressed and secreted in cell culture as described in Example 2. The isolated protein was then sequenced at its N-terminal region, yielding the N-terminal sequence shown in Fig. 8. This sequence, which is identified herein as SEQ ID NO:22, has the same N-terminal residues as native mature α_1 -antitrypsin.

10 Production of mature ATIII: In a preferred embodiment, the culture medium contains a MES buffer, pH 6.8. Western blot analysis of the ATIII-protein produced, shown in lanes 4 and 6 in Fig. 9, shows a band corresponding to ATIII (lane 1) in cell lines 42 and 46, when grown in the absence (but not in the presence) of sucrose.

15 Production of mature BPN': In one embodiment of the invention, in which BPN' is secreted as the proBPN' form of the enzyme, the chaperon "pro" moiety of the enzyme facilitates enzyme folding and is cleaved from the enzyme, leaving the active mature form of BPN'. In another embodiment, the mature enzyme is co-expressed and co-secreted with the "pro" chaperon moiety, with conversion of the enzyme to active form occurring in presence of the free chaperon (Eder *et al.*, *Biochem.* (1993) 32:18-26; Eder *et al.*, (1993) *J. Mol. Biol.* 223:293-304). In yet another 20 embodiment of the invention, the BPN' is secreted in inactive form at a pH that may be in the 6-8 range, with subsequent activation of the inactive form, *e.g.*, after enzyme isolation, by exposure to the "pro" chaperon moiety, *e.g.*, immobilized to a solid support.

25 In both of these embodiments, the culture medium is maintained at a pH of between 5 and 6, preferably about 5.5 during the period of active expression and secretion of BPN', to keep the BPN', which is normally active at alkaline pH, at a pH below optimal activity.

30 Codon optimization to the host plant's most frequent codons yielded a severalfold enhancement in the level of expressed heterologous protein in cell culture as shown in Fig. 11. The extent of enhancement is seen from the Western blot analysis shown in Fig. 10 for two cells lines and further substantiated in Fig. 11. Lane 2 (second from left) in Fig. 10 shows a Western blot of BPN' obtained in culture from cells transformed with a native proBPN' coding sequence. Two bands observed correspond to a lower molecular weight protein whose approximately 35 kdal molecular weight corresponds to that of proBPN'. The upper band corresponds to a somewhat higher molecular weight species, possibly glycosylated.

35 The first lane in the figure shows BPN' polypeptides produced in culture by plant cells transformed with the codon-optimized proBPN' sequence identified by SEQ ID NO:21. For

comparative purposes, the same volume of culture medium, adjusted for cell density, was applied in both lanes 1 and 2. As seen, the amount of BPN' enzyme produced with a codon-optimized sequence was severalfold higher than for subtilisin BPN' produced with the native coding sequence. Further, a dark band or bands corresponding to mature peptide (molecular weight 27.5 kdal) was 5 observed. However, it should be noted that directly above the band at 35kD is a more pronounced band which may be pro mature product yet to be cleaved into active form.

Fig. 11 compares the specific activity of BPN' codon-optimized (AP106) versus BPN' native (AP101) expression in rice callus cell culture, assayed using the chromogenic peptide substrate suc-Ala-Ala-Pro-Phe-pNA as described by DelMar, E.G. *et al.* (1979; *Anal. Biochem.* 10 99:316-320). As shown in Fig. 11, several of the cell lines transformed with codon-optimized chimeric genes produced levels of BPN', as evidenced by measured specific activity in culture medium, that were 2-5 times the highest levels observed for plant cells transformed with native proBPN' sequence.

In accordance with another aspect of the invention, it has been found that the transformed 15 plant cell culture is able to express and secrete BPN' at a cell culture pH, pH 5.5, which largely inhibits self-degradation of mature, active BPN'. To assay for optimal pH conditions, the assay disclosed in DelMar, *et al.* (*supra*) is used to test the media derived from BPN' transformed cell lines under various pH conditions. Transformed rice callus cells are cultured in a MES medium 20 under similar conditions as disclosed in Example 2, but where the pH of the medium is maintained at a selected pH between 5 and 8.0. At each pH, the total amount of expressed and secreted BPN' is determined by Western blot analysis. BPN' activity can be tested in the assay described by DelMar (*supra*).

V. Production of Mature Heterologous Protein in Germinating Seeds

25 In this embodiment, monocot cells transformed as above are used to regenerate plants, seeds from the plants are harvested and then germinated, and the mature protein is isolated from the germinated seeds.

Plant regeneration from cultured protoplasts or callus tissue is carried by standard methods, *e.g.*, as described in Evans *et al.*, HANDBOOK OF PLANT CELL CULTURES Vol. 1: (MacMillan 30 Publishing Co. New York, 1983); and Vasil I.R. (ed.), CELL CULTURE AND SOMATIC CELL GENETICS OF PLANTS, Acad. Press, Orlando, Vol. I, 1984, and Vol. III, 1986, and as described in the above-cited PCT application.

A. Seed Germination Conditions

The transgenic seeds obtained from the regenerated plants are harvested, and prepared for 35 germination by an initial steeping step, in which the seeds immersed in or sprayed with water to

increase the moisture content of the seed to between 35-45%. This initiates germination. Steeping typically takes place in a steep tank which is typically fitted with a conical end to allow the seed to flow freely out. The addition of compressed air to oxygenate the steeping process is an option. The temperature is controlled at approximately 22°C depending on the seed.

5 After steeping, the seeds are transferred to a germination compartment which contains air saturated with water and is under controlled temperature and air flows. The typical temperatures are between 12-25°C and germination is permitted to continue for from 3 to 7 days.

Where the heterologous protein coding gene is operably linked to a inducible promoter requiring a metabolite such as sugar or plant hormone, *e.g.*, 2 to 100 μ M gibberellic acid, this 10 metabolite is added, removed or depleted from the steeping water medium and/or is added to the water saturated air used during germination. The seed absorbs the aqueous medium and begins to germinate, expressing the heterologous protein. The medium may then be withdrawn and the malting begun, by maintaining the seeds in a moist temperature controlled aerated environment. In this way, the seeds may begin growth prior to expression, so that the expressed product is less 15 likely to be partially degraded or denatured during the process.

More specifically, the temperature during the imbibition or steeping phase will be maintained in the range of about 15-25°C, while the temperature during the germination will usually be about 20°C. The time for the imbibition will usually be from about 1 to 4 days, while the germination time will usually be an additional 1 to 10 days, more usually 3 to 7 days. Usually, the 20 time for the malting does not exceed about ten days. The period for the malting can be reduced by using plant hormones during the imbibition, particularly gibberellic acid.

To achieve maximum production of recombinant protein from malting, the malting procedure may be modified to accommodate de-hulled and de-embryonated seeds, as described in above-cited PCT application WO 95/14099. In the absence of sugars from the endosperm, there is 25 expected to be a 5 to 10 fold increase in RAmy3D promoter activity and thus expression of heterologous protein. Alternatively when embryoless half-seeds are incubated in 10 mM CaCl₂ and 5 μ M gibberellic acid, there is a 50 fold increase in RAmy1A promoter activity.

Production of mature HSA: Following the germination conditions as outlined above and further detailed in Example 3, supernatant was analyzed by Western blot. Western blot analysis 30 shows production of HSA in germinating rice seeds, with seed samples taken 24, 72, and 120 hours after induction with gibberellin. HSA production was highest approximately 24 hours post-induction (lanes 3 and 4, Fig. 12). Bilirubin binding, a measure of correct folding of plant-produced HSA, is assayed according to the method presented in Example 3.

35 VI. Production of Mature Heterologous Protein in Maturing Seeds

In this embodiment, monocot cells transformed as above are used to regenerate plants, and seeds from the plants are allowed to mature, typically in the field, with consequent production of heterologous protein in the seeds.

Following seed maturation, the seeds and their heterologous proteins may be used directly, 5 that is, without protein isolation, where for example, the heterologous protein is intended to confer a benefit on the seed as a whole, for example, to enrich the seed in the selected protein.

Alternatively, the seeds may be fractionated by standard methods to obtain the heterologous protein in enriched or purified form. In one general approach, the seed is first milled, then suspended in a suitable extraction medium, e.g., an aqueous or an organic solvent, to extract the 10 protein or metabolite of interest. If desired the heterologous protein can be further fractionated and purified, using standard purification methods.

The following examples are provided by way of illustration only and not by way of limitation. Those of skill will readily recognize a variety of noncritical parameters which could be 15 changed or modified to yield essentially similar results.

General Methods

Generally, the nomenclature and laboratory procedures with respect to standard recombinant DNA technology can be found in Sambrook, *et al.*, MOLECULAR CLONING - A LABORATORY 20 MANUAL, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 1989 and in S.B. Gelvin and R.A. Schilperoort, PLANT MOLECULAR BIOLOGY, 1988. Other general references are provided throughout this document. The procedures therein are known in the art and are provided for the convenience of the reader.

25

Example 1

Construction of a Transforming Vector Containing a Codon-Optimized α_1 -antitrypsin Sequence

A. Hygromycin Resistance Gene Insertion:

The 3 kb *Bam*HI fragment containing the 35S promoter-Hph-NOS was removed from the 30 plasmid pMON410 (Monsanto, St. Louis, MO) and placed into an site-directed mutagenized *Bgl*II site in the pUC18 at 1463 to form the plasmid pUCH18+.

B. Terminator Insertion:

pOSg1ABK5 is a 5 kb *Bam*HI-*Kpn*I fragment from lambda clone λ OSg1A (Huang, N., *et 35 al.*, (1990) Nuc. Acids Res. 18:7007) cloned into pBluescript KS- (Stratagene, San Diego, CA).

Plasmid pOSg1ABK5 was digested with *Msp*I and blunted with T4 DNA polymerase followed by *Spe*I digestion. The 350 bp terminator fragment was subcloned into pUC19 (New England BioLabs, Beverly, MA), which had been digested with *Bam*HI, blunted with T4 DNA polymerase and digested with *Xba*I, to form pUC19/terminator.

5

C. RAmy3D Promoter Insertion:

A 1.1 kb *Nhe*I-*Pst*I fragment derived from p1AS1.5 (Huang, N. *et al.* (1993) Plant Mol. Biol. 23:737-747), was cloned into the vector pGEM5zf- [multiple cloning site (MCS) (Promega, Madison, WI): *Ap*AI, *Aat*II, *Sph*I, *Nco*I, *Sst*II, *Eco*RV, *Spe*I, *Not*I, *Pst*I, *Sal*I, *Nde*I, *Sac*I, *Mlu*I, *Nsi*I] at the *Spe*I and *Pst*I sites to form pGEM5zf-(3D/*Nhe*I-*Pst*I). pGEM5zf-(3D/*Nhe*I-*Pst*I) was then digested with *Pst*I and *Sac*I, and two non-kinased 30mers having the complementary sequences 5' GCTTG ACCTG TAACT CGGGC CAGGC GAGCT 3' (SEQ ID NO:23) and 5' CGCCT AGCCC GAGTT ACAGG TCAAG CAGCT 3' (SEQ ID NO:24) were ligated in to form p3DProSig. The promoter fragment prepared by digesting p3DProSig with *Nco*I, blunting with T4 DNA polymerase, and digesting with *Sst*I was subcloned into pUC19/terminator which had been digested with *Eco*RI, blunted with T4 DNA polymerase and digested with *Sst*I, to form p3DProSigEND.

D. Multiple Cloning Site Insertion:

p3DProSigEND was digested with *Sst*I and *Sma*I followed by the ligation of a new synthetic linker fragment constructed with the non-kinased complementary oligonucleotides 5' AGCTC CATGG CCGTG GCTCG AGTCT AGACG CGTCC CC 3' (SEQ ID NO:25) and 5' GGGGA CGCGT CTAGA CTCGA GCCAC GGCCA TGG 3' (SEQ ID NO:26) to form p3DProSigENDlink.

25

E. p3DProSigENDlink Flanking Site Modification:

p3DProSigENDlink was digested with *Sal*I and blunted with T4 DNA polymerase followed by *Eco*RV digestion. The blunt fragment was then inserted into pBluescript KS+ (Stratagene) in the *Eco*RV site so that the *Hind*III site is proximal to the promoter and the *Eco*RI is proximal to the terminator sequence. The *Hind*III-*Eco*RI fragment was then moved into the polylinker of pUCH18+ to form the p3Dv1.0 expression vector.

F. RAmy1A Promoter Insertion:

A 1.9 kb *Nhe*I-*Pst*I fragment derived from subclone pOSG2CA2.3 from lambda clone λ OSg2 (Huang *et al.* (1990) Plant Mol. Biol. 14:655-668), was cloned into the vector pGEM5zf- at

the *SpeI* and *PstI* sites to form pGEM5zf-(1A/*NheI-PstI*). pGEM5zf-(1A/*NheI-PstI*) was digested with *PstI* and *SacI* and two non-kinased 35mers and four kinased 32mers were ligated in, with the complementary sequences as follows: 5' GCATG CAGGT GCTGA ACACC ATGGT GAACA AACAC 3' (SEQ ID NO:27); 5' TTCTT GTCCC TTTCG GTCCT CATCG TCCTC CT 3' (SEQ ID NO:28); 5' TGGCC TCTCC TCCAA CTTGA CAGCC GGGAG CT 3' (SEQ ID O:29); 5' TTCAC CATGG TGTTC AGCAC CTGCA TGCTG CA 3' (SEQ ID NO:30); 5' CGATG AGGAC CGAAA GGGAC AAGAA GTGTT TG 3' (SEQ ID NO:31); 5' CCCGG CTGTC AAGTT GGAGG AGAGG CCAAG GAGGA 3' (SEQ ID NO:32) to form p1AProSig. The *HindIII-SacI* 0.8 kb promoter fragment was subcloned from p1AProSig into the p3Dv1.0 vector digested with *HindIII-SacI* to yield the p1Av1.0 expression vector.

G. Construction of p3D-AAT Plasmid

Two PCR primers were used to amplify a fragment encoding AAT according to the sequence disclosed as Genbank Accession No. K01396: N-terminal primer 5' GAGGA TCCCC AGGGA GATGC TGCCC AGAA 3' (SEQ ID NO:33) and C-terminal primer 5' CGCGC TCGAG TTATT TTTGG GTGGG ATTCA CCAC 3' (SEQ ID NO:34). The N-terminal primer amplifies to a blunt site for in-frame insertion with the end of the p3D signal peptide and the C-terminal primer contains a *XhoI* site for cloning the fragment into the vector as shown in Figs. 3A and 3B. Alternatively, the sequence encoding mature AAT (SEQ ID NO:8) or codon-optimized AAT may be chemically synthesized using techniques known in the art, incorporating a *XhoI* restriction site 3' of the termination codon for insertion into the expression vector as described above.

Example 2

Production of mature α -antitrypsin in cell culture

25 After selection of transgenic callus, callus cells were suspended in liquid culture containing AA2 media (Thompson, J.A., *et al.*, *Plant Science* 47:123 (1986), at 3% sucrose, pH 5.8. Thereafter, the cells were shifted to phosphate-buffered media (20 mM phosphate buffer, pH 6.8) using 10 mL multi-well tissue culture plates and shaken at 120 rpm in the dark for 48 hours. The supernatant was then removed and stored at -80°C prior to western blot analysis.

30 Supernatants were concentrated using Centricon-10 filters (Amicon cat. #4207) and washed with induction media to remove substances interfering with electrophoretic migration. Samples were concentrated approximately 10 fold, and mature AAT was purified by SDS PAGE electrophoresis. The purified protein was extracted from the electrophoresis medium, and sequenced at its N-terminus, giving the sequence shown in Fig. 8, identified herein as SEQ ID NO:22.

Example 3HSA Induction in Germinating Seeds

5 After selection of transgenic plants which tested positive for the presence of a codon-optimized HSA gene driven by the GA₃-responsive RAmy1A promoter, seeds were harvested and imbibed for 24 hours with 100 rpm orbital shaking in the dark at 25°C. GA₃ was added to a final concentration of 5μM and incubated for an additional 24-120 hours. Total soluble protein was isolated by double grinding each seed in 120 μl grinding buffer and centrifuging at 23,000 x g for 1
10 minute at 4°C. The clear supernatant was carefully removed from the pellet and transferred to a fresh tube.

Bilirubin binding assay

15 Bilirubin binding to its high-affinity site on mature HSA is assayed using the method described by Jacobsen, J. *et al.* (1974; Clin. Chem. 20:783) and Reed, R.G. *et al.* (1975; Biochemistry 14:4578-4583). Briefly, the concentration of free bilirubin in equilibrium with protein-bound bilirubin is determined by the rate of peroxide-peroxidase catalyzed oxidation of free bilirubin. Stock solutions of bilirubin (Nutritional Biochemicals Corp.) are prepared fresh daily in 5 mM NaOH containing 1mM EDTA and the concentration determined using a molar absorptivity of 47,500 M⁻¹ cm⁻¹ at 440 nm. An aliquot containing between 5 and 30 nmol bilirubin is added to a
20 1 cm cuvette containing 1 ml PBS and approximately 30 nmol HSA at 37°C. An absorbance spectrum between 500 and 350 nm is recorded. Aliquots of horseradish peroxidase (Sigma), 0.05 mg/ml in PBS, and 0.05% ethyl hydrogen peroxide (Ferrosan; Malmö Sweden) are added and the change in absorbance at λ max is recorded for 3-5 minutes. The concentrations of free and bound
25 bilirubin calculated from the oxidation rate observed using varying concentrations of total bilirubin are used to construct a Scatchard plot from which the association constant for a single binding site is determined.

30 Although the invention has been described with reference to particular embodiments, it will be appreciated that a variety of changes and modifications can be made without departing from the invention.

35

SEQUENCE LISTING

5 (1) GENERAL INFORMATION

(i) APPLICANT: Applied Phytologics, Inc.

10 (ii) TITLE OF THE INVENTION: Production of Mature Proteins
in Plants

(iii) NUMBER OF SEQUENCES: 34

15 (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Dehlinger & Associates
(B) STREET: P.O. Box 60850
(C) CITY: Palo Alto
(D) STATE: CA
(E) COUNTRY: USA
20 (F) ZIP: 94306

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
25 (C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US98/03068
30 (B) FILING DATE: 13-FEB-1998
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 60/038,169
35 (B) FILING DATE: 13-FEB-1997

(A) APPLICATION NUMBER: 60/037,991
(B) FILING DATE: 13-FEB-1997

40 (A) APPLICATION NUMBER: 60/038,170
(B) FILING DATE: 13-FEB-1997

(A) APPLICATION NUMBER: 60/038,168
(B) FILING DATE: 13-FEB-1997

45 (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Petithory, Joanne R
(B) REGISTRATION NUMBER: P42,995
(C) REFERENCE/DOCKET NUMBER: 0665-0007.41

50 (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 650-324-0880
(B) TELEFAX: 650-324-0960

55 (2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
60 (B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:
65 (B) CLONE: 3D signal peptide sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met	Lys	Asn	Thr	Ser	Ser	Leu	Cys	Leu	Leu	Leu	Val	Val	Leu	Cys
1				5				10					15	
Ser	Leu	Thr	Cys	Asn	Ser	Gly	Gln	Ala						
				20				25						

5

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 75 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (B) CLONE: native 3D signal peptide DNA sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGAAGAACAA	CCAGCAGCTT	GTGTTTGTCT	CTCCTCGTGG	TGCTCTGCAG	CTTGACCTGT	60
AACTCGGGCC	AGGCG					75

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 75 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (B) CLONE: codon-optimized 3D signal peptide DNA sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGAAGAACAA	CCTCCTCCCT	CTGCCTCCTG	CTGCTCGTGG	TCCTCTGCTC	CCTGACCTGC	60
AACAGCGGCC	AGGCG					75

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (vii) IMMEDIATE SOURCE:
- (B) CLONE: RAmy1A signal peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Val	Asn	Lys	His	Phe	Leu	Ser	Leu	Ser	Val	Leu	Ile	Val	Leu	Leu
1					5					10			15		
Gly	Leu	Ser	Ser	Asn	Leu	Thr	Ala	Gly							
					20			25							

55

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (B) CLONE: RAmy 1A 5' untranslated region (UTR)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATCAATCATC CATCTCCGAA GTGTGTCTGC AGCATGCAGG TGCTAACAC C

51

(2) INFORMATION FOR SEQ ID NO:6:

5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 321 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10 (vii) IMMEDIATE SOURCE:
 (B) CLONE: RAmy 1A 3' untranslated region (UTR)

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

13	GCGCACGATG ACGAGACTCT CAGTTAGCA GATTAAACCT GCGATTTTA CCCTGACCGG	60
	TATACGTATA TACGTGCCGG CAACGAGCTG TATCCGATCC GAATTACGGA TGCAATTGTC	120
	CACGAAGTAC TTCCTCCGTA AATAAAGTAG GATCAGGGAC ATACATTGAT ATGGTTTTAC	180
18	GAATAATGCT ATGCAATAAA ATTTGCACTG CTTAATGCTT ATGCATTTTT GCTTGGTTCG	240
20	ATTGTACTGG TGAATTATTG TTACTGTTCT TTTTACTTCT CGAGTGGCAG TATTGTTCTT	300
	CTACGAAAAT TTGATGCGTA G	321

(2) INFORMATION FOR SEQ ID NO:7:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 394 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: protein
 (vii) IMMEDIATE SOURCE:
 (B) CLONE: mature AAT amino acid sequence

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

13	Glu Asp Pro Gln Gly Asp Ala Ala Gln Lys Thr Asp Thr Ser His His	
	1 5 10 15	
	Asp Gln Asp His Pro Thr Phe Asn Lys Ile Thr Pro Asn Leu Ala Glu	
	20 25 30	
18	Phe Ala Phe Ser Leu Tyr Arg Gln Leu Ala His Gln Ser Asn Ser Thr	
	35 40 45	
	Asn Ile Phe Phe Ser Pro Val Ser Ile Ala Thr Ala Phe Ala Met Leu	
	50 55 60	
	Ser Leu Gly Thr Lys Ala Asp Thr His Asp Glu Ile Leu Glu Gly Leu	
23	65 70 75 80	
	Asn Phe Asn Leu Thr Glu Ile Pro Glu Ala Gln Ile His Glu Gly Phe	
	85 90 95	
	Gln Glu Leu Leu Arg Thr Leu Asn Gln Pro Asp Ser Gln Leu Gln Leu	
	100 105 110	
28	Thr Thr Gly Asn Gly Leu Phe Leu Ser Glu Gly Leu Lys Leu Val Asp	
	115 120 125	
	Lys Phe Leu Glu Asp Val Lys Lys Leu Tyr His Ser Glu Ala Phe Thr	
	130 135 140	
	Val Asn Phe Gly Asp Thr Glu Glu Ala Lys Lys Gln Ile Asn Asp Tyr	
33	145 150 155 160	
	Val Glu Lys Gly Thr Gln Gly Lys Ile Val Asp Leu Val Lys Glu Leu	
	165 170 175	
	Asp Arg Asp Thr Val Phe Ala Leu Val Asn Tyr Ile Phe Phe Lys Gly	
	180 185 190	
38	Lys Trp Glu Arg Pro Phe Glu Val Lys Asp Thr Glu Glu Asp Phe	
	195 200 205	
	His Val Asp Gln Val Thr Thr Val Lys Val Pro Met Met Lys Arg Leu	
	210 215 220	
	Gly Met Phe Asn Ile Gln His Cys Lys Lys Leu Ser Ser Trp Val Leu	
43	225 230 235 240	
	Leu Met Lys Tyr Leu Gly Asn Ala Thr Ala Ile Phe Phe Leu Pro Asp	
	245 250 255	

	Glu	Gly	Lys	Leu	Gln	His	Leu	Glu	Asn	Glu	Leu	Thr	His	Asp	Ile	Ile
	260							265						270		
	Thr	Lys	Phe	Leu	Glu	Asn	Glu	Asp	Arg	Arg	Ser	Ala	Ser	Leu	His	Leu
	275							280						285		
5	Pro	Lys	Leu	Ser	Ile	Thr	Gly	Thr	Tyr	Asp	Leu	Lys	Ser	Val	Leu	Gly
	290							295						300		
	Gln	Leu	Gly	Ile	Thr	Lys	Val	Phe	Ser	Asn	Gly	Ala	Asp	Leu	Ser	Gly
	305							310						315		
10	Val	Thr	Glu	Glu	Ala	Pro	Leu	Lys	Leu	Ser	Lys	Ala	Val	His	Lys	Ala
	325							330						335		
	Val	Leu	Thr	Ile	Asp	Glu	Lys	Gly	Thr	Glu	Ala	Ala	Gly	Ala	Met	Phe
	340							345						350		
	Leu	Glu	Ala	Ile	Pro	Met	Ser	Ile	Pro	Pro	Glu	Val	Lys	Phe	Asn	Lys
	355							360						365		
15	Pro	Phe	Val	Phe	Leu	Met	Ile	Glu	Gln	Asn	Thr	Lys	Ser	Pro	Leu	Phe
	370							375						380		
	Met	Gly	Lys	Val	Val	Asn	Pro	Thr	Gln	Lys						
	385							390								

20 (2) INFORMATION FOR SEQ ID NO:8:

	(i) SEQUENCE CHARACTERISTICS:															
	(A) LENGTH: 1185 base pairs															
	(B) TYPE: nucleic acid															
25	(C) STRANDEDNESS: single															
	(D) TOPOLOGY: linear															
	(vii) IMMEDIATE SOURCE:															
	(B) CLONE: native coding sequence of mature AAT															
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:															
	GAGGATCCCC	AGGGAGATGC	TGCCAGAGAAG	ACAGATACAT	CCCACCATGA	TCAGGATCAC										60
35	CCAACCTTCA	ACAAGATCAC	CCCCAACCTG	GCTGAGTTCG	CCTTCAGCCT	ATACCGCCAG										120
	CTGGCACACC	AGTCCAACAG	CACCAATATC	TTCTTCTCCC	CAGTGAGCAT	CGCTACAGCC										180
	TTTGCAATGC	TCTCCCTGGG	GACCAAGGCT	GACACTCACG	ATGAAATCCT	GGAGGGCCTG										240
	AATTTCAAACC	TCACGGAGAT	TCCGGAGGCT	CAGATCCATG	AAGGCTTCCA	GGAACCTCTC										300
	CGTACCTCTA	ACCAGCCAGA	CAGCCAGCTC	CAGCTGACCA	CCGGCAATGG	CCTGTTCCCTC										360
40	AGCGAGGGCC	TGAAGCTAGT	GGATAAGTTT	TTGGAGGATG	TTAAAAAGTT	GTACCACTCA										420
	GAAGCCTTCA	CTGTCAACTT	CGGGGACACC	GAAGAGGCCA	AGAAACAGAT	CAACGATTAC										480
	GTGGAGAAGG	GTACTCAAGG	GAAAATTGTG	GATTTGGTCA	AGGAGCTTGA	CAGAGACACA										540
	GTTTTTGCTC	TGGTGAATTA	CATCTCTTT	AAAGGCAAAT	GGGAGAGACC	CTTTGAAGTC										600
	AAGGACACCG	AGGAAGAGGA	CTTCCACGTG	GACCAGGTGA	CCACCGTGAA	GGTGCCTATG										660
45	ATGAAGCGTT	TAGGCATGTT	TAACATCCAG	CACTGTAAGA	AGCTGTCAG	CTGGGTGCTG										720
	CTGATGAAAAT	ACCTGGCAA	TGCCACCGCC	ATCTTCTTCC	TGCCTGATGA	GGGGAAACTA										780
	CAGCACCTGG	AAAATGAACT	CACCCACGAT	ATCATCACCA	AGTTCTTGGA	AAATGAAGAC										840
	AGAAGGTCTG	CCAGCTTACA	TTTACCCAAA	CTGTCCATT	CTGGAACCTA	TGATCTGAAG										900
	AGCGTCTGG	GTCAACTGGG	CATCACTAAG	GTCTTCAGCA	ATGGGGCTGA	CCTCTCCGGG										960
50	GTCACAGAGG	AGGCACCCCT	GAAGCTCTCC	AAGGCCGTGC	ATAAGGCTGT	GCTGACCATC										1020
	GACGAGAAAG	GGACTGAAGC	TGCTGGGCC	ATGTTTTAG	AGGCCATACC	CATGTCTATC										1080
	CCCCCGAGG	TCAAGTTCAA	CAAACCTTT	GTCTTCTAA	TGATTGAACA	AAATACCAAG										1140
	TCTCCCCTCT	TCATGGGAAA	AGTGGTGAAT	CCCACCCAAA	AATAA											1185

55 (2) INFORMATION FOR SEQ ID NO:9:

	(i) SEQUENCE CHARACTERISTICS:														
	(A) LENGTH: 432 amino acids														
	(B) TYPE: amino acid														
60	(D) TOPOLOGY: linear														
	(ii) MOLECULE TYPE: protein														
	(vii) IMMEDIATE SOURCE:														
	(B) CLONE: mature ATIII aa sequence														
65	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:														

His Gly Ser Pro Val Asp Ile Cys Thr Ala Lys Pro Arg Asp Ile Pro

	1	5	10	15
	Met Asn Pro Met Cys Ile Tyr Arg Ser Pro Glu Lys Lys Ala Thr Glu			
	20 25 30			
5	Asp Glu Gly Ser Glu Gln Lys Ile Pro Glu Ala Thr Asn Arg Arg Val			
	35 40 45			
	Trp Glu Leu Ser Lys Ala Asn Ser Arg Phe Ala Thr Thr Phe Tyr Gln			
	50 55 60			
10	His Leu Ala Asp Ser Lys Asn Asp Asn Asn Ile Phe Leu Ser Pro			
	65 70 75 80			
	Leu Ser Ile Ser Thr Ala Phe Ala Met Thr Lys Leu Gly Ala Cys Asn			
	85 90 95			
	Asp Thr Leu Gln Gln Leu Met Glu Val Phe Lys Phe Asp Thr Ile Ser			
	100 105 110			
15	Glu Lys Thr Ser Asp Gln Ile His Phe Phe Ala Lys Leu Asn Cys			
	115 120 125			
	Arg Leu Tyr Arg Lys Ala Asn Lys Ser Ser Lys Leu Val Ser Ala Asn			
	130 135 140			
20	Arg Leu Phe Gly Asp Lys Ser Leu Thr Phe Asn Glu Thr Tyr Gln Asp			
	145 150 155 160			
	Ile Ser Glu Leu Val Tyr Gly Ala Lys Leu Gln Pro Leu Asp Phe Lys			
	165 170 175			
	Glu Asn Ala Glu Gln Ser Arg Ala Ala Ile Asn Lys Trp Val Ser Asn			
	180 185 190			
25	Lys Thr Glu Gly Arg Ile Thr Asp Val Ile Pro Ser Glu Ala Ile Asn			
	195 200 205			
	Glu Leu Thr Val Leu Val Leu Val Asn Thr Ile Tyr Phe Lys Gly Leu			
	210 215 220			
30	Trp Lys Ser Lys Phe Ser Pro Glu Asn Thr Arg Lys Glu Leu Phe Tyr			
	225 230 235 240			
	Lys Ala Asp Gly Glu Ser Cys Ser Ala Ser Met Met Tyr Gln Glu Gly			
	245 250 255			
	Lys Phe Arg Tyr Arg Arg Val Ala Glu Gly Thr Gln Val Leu Glu Leu			
	260 265 270			
35	Pro Phe Lys Gly Asp Asp Ile Thr Met Val Leu Ile Leu Pro Lys Pro			
	275 280 285			
	Glu Lys Ser Leu Ala Lys Val Glu Lys Glu Leu Thr Pro Glu Val Leu			
	290 295 300			
40	Gln Glu Trp Leu Asp Glu Leu Glu Glu Met Met Leu Val Val His Met			
	305 310 315 320			
	Pro Arg Phe Arg Ile Glu Asp Gly Phe Ser Leu Lys Glu Gln Leu Gln			
	325 330 335			
	Asp Met Gly Leu Val Asp Leu Phe Ser Pro Glu Lys Ser Lys Leu Pro			
	340 345 350			
45	Gly Ile Val Ala Glu Gly Arg Asp Asp Leu Tyr Val Ser Asp Ala Phe			
	355 360 365			
	His Lys Ala Phe Leu Glu Val Asn Glu Glu Gly Ser Glu Ala Ala Ala			
	370 375 380			
50	Ser Thr Ala Val Val Ile Ala Gly Arg Ser Leu Asn Pro Asn Arg Val			
	385 390 395 400			
	Thr Phe Lys Ala Asn Arg Pro Phe Leu Val Phe Ile Arg Glu Val Pro			
	405 410 415			
	Leu Asn Thr Ile Ile Phe Met Gly Arg Val Ala Asn Pro Cys Val Lys			
55	420 425 430			

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1299 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (B) CLONE: native ATIII DNA sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

	CACGGAAAGCC	CTGTGGACAT	CTGCACAGCC	AAGCCGCGGG	ACATCCCAT	GAATCCCATG	60
	TGCATTTACC	GCTCCCCGGA	GAAGAAGGCA	ACTGAGGATG	AGGGCTCAGA	ACAGAAGATC	120
	CCGGAGGCCA	CCAACCGGCG	TGTCTGGAA	CTGTCCAAGG	CCAATCCCC	CTTTGCTACC	180
5	ACTTTCTATC	AGCACCTGGC	AGATTCCAAG	AATGACAATG	ATAACATTTT	CCTGTACCC	240
	CTGAGTATCT	CCACGGCTT	TGCTATGACC	AAGCTGGGTG	CCTGTAATGA	CACCCCTCAG	300
	CAACTGATGG	AGGTATTTAA	TTTGACACC	ATATCTGAGA	AAACATCTGA	TCAGATCCAC	360
	TTCTTCTTG	CCAAACTGAA	CTGCCGACTC	TATCGAAAAG	CCAACAAATC	CTCCAAGTTA	420
	GTATCAGCCA	ATCGCCTTT	TGGAGACAAA	TCCCTTACCT	TCAATGAGAC	CTACCAGGAC	480
	ATCAGTGAGT	TGGTATATGG	AGCCAAGCTC	CAGCCCCTGG	ACTTCAAAGGA	AAATGCAGAG	540
10	CAATCCAGAG	CGGCCATCAA	CAAATGGGTG	TCCAATAAGA	CCGAAGGCG	AATCACCGAT	600
	GTCAAGGGCC	CGGAAGCCAT	CAATGAGCTC	ACTGTTCTGG	TGCTGTTAA	CACCATTTAC	660
	TGTTGGAAGTC	AAAGTTCAGC	CCTGAGAACAA	CAAGGAAGGA	ACTGTTCTAC	720	
	AAGGCTGATG	GAGAGTCGTG	TTCAGCATCT	ATGATGTACC	AGGAAGGCAA	GTTCCGTTAT	780
	CGGCGCGTGG	CTGAAGGCAC	CCAGGTGCTT	GAGTTGCCCT	TCAAAGGTGA	TGACATCACC	840
15	ATGGTCCTCA	TCTTGCCCAA	GCCTGAGAAG	AGCCTGGCCA	AGGTGGAGAA	GGAACCTCACC	900
	CCAGAGGTGC	TGCAGGAGTG	GCTGGATGAA	TTGGAGGAGA	TGATGCTGGT	GGTTCACATG	960
	CCCCGCTTCC	GCATTGAGGA	CGGCTTCAGT	TTGAAGGAGC	AGCTGCAAGA	CATGGGCCTT	1020
	GTCGATCTGT	TCAGCCCTGA	AAAGTCCAAA	CTCCCAGGTA	TTGTTGCAGA	AGGCGAGAT	1080
20	GACCTCTATG	TCTCAGATGC	ATTCCATAAG	GCATTCTTG	AGGTAAATGA	AGAAGGCAGT	1140
	GAAGCAGCTG	CAAGTACCGC	TGTTGTGATT	GCTGGCCGTT	CGCTAAACCC	CAACAGGGTG	1200
	ACTTTCAAGG	CCAACAGGCC	CTTCCTGGTT	TTTATAAGAG	AAGTTCCCTCT	GAACACTATT	1260
	ATCTTCATGG	GCAGAGTAGC	CAACCCTTGT	GTAAAGTAA			1299

(2) INFORMATION FOR SEQ ID NO:11:

25

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 585 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

30

(ii) MOLECULE TYPE: protein
 (vii) IMMEDIATE SOURCE:
 (B) CLONE: mature HSA amino acid sequence

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

	Asp	Ala	His	Lys	Ser	Glu	Val	Ala	His	Arg	Phe	Lys	Asp	Leu	Gly	Glu
	1					5			10		15					
	Glu	Asn	Phe	Lys	Ala	Leu	Val	Leu	Ile	Ala	Phe	Ala	Gln	Tyr	Leu	Gln
40						20			25		30					
	Gln	Cys	Pro	Phe	Glu	Asp	His	Val	Lys	Leu	Val	Asn	Glu	Val	Thr	Glu
						35			40		45					
	Phe	Ala	Lys	Thr	Cys	Val	Ala	Asp	Glu	Ser	Ala	Glu	Asn	Cys	Asp	Lys
45						50			55		60					
	Ser	Leu	His	Thr	Leu	Phe	Gly	Asp	Lys	Leu	Cys	Thr	Val	Ala	Thr	Leu
50						65			70		75		80			
	Arg	Glu	Thr	Tyr	Gly	Glu	Met	Ala	Asp	Cys	Cys	Ala	Lys	Gln	Glu	Pro
55						85			90		95					
	Glu	Arg	Asn	Glu	Cys	Phe	Leu	Gln	His	Lys	Asp	Asp	Asn	Pro	Asn	Leu
60						100			105		110					
	Pro	Arg	Leu	Val	Arg	Pro	Glu	Val	Asp	Val	Met	Cys	Thr	Ala	Phe	His
65						115			120		125					
	Asp	Asn	Glu	Glu	Thr	Phe	Leu	Lys	Lys	Tyr	Leu	Tyr	Glu	Ile	Ala	Arg
70						130			135		140					
	Arg	His	Pro	Tyr	Phe	Tyr	Ala	Pro	Glu	Leu	Leu	Phe	Phe	Ala	Lys	Arg
75						145			150		155		160			
	Tyr	Lys	Ala	Ala	Phe	Thr	Glu	Cys	Cys	Gln	Ala	Ala	Asp	Lys	Ala	Ala
80						165			170		175					
	Cys	Leu	Leu	Pro	Lys	Leu	Asp	Glu	Leu	Arg	Asp	Glu	Gly	Lys	Ala	Ser
85						180			185		190					
	Ser	Ala	Lys	Gln	Arg	Leu	Lys	Cys	Ala	Ser	Leu	Gln	Lys	Phe	Gly	Glu
90						195			200		205					
	Arg	Ala	Phe	Lys	Ala	Trp	Ala	Val	Ala	Arg	Leu	Ser	Gln	Arg	Phe	Pro
95						210			215		220					
	Lys	Ala	Glu	Phe	Ala	Glu	Val	Ser	Lys	Leu	Val	Thr	Asp	Leu	Thr	Lys
100						225			230		235		240			
	Val	His	Thr	Glu	Cys	Cys	His	Gly	Asp	Leu	Leu	Glu	Cys	Ala	Asp	Asp

	245	250	255
	Arg Ala Asp Leu Ala Lys Tyr Ile Cys Glu Asn Gln Asp Ser Ile Ser		
	260	265	270
5	Ser Lys Leu Lys Glu Cys Cys Glu Lys Pro Leu Leu Glu Lys Ser His		
	275	280	285
	Cys Ile Ala Glu Val Glu Asn Asp Glu Met Pro Ala Asp Leu Pro Ser		
	290	295	300
	Leu Ala Ala Asp Phe Val Glu Ser Lys Asp Val Cys Lys Asn Tyr Ala		
	305	310	315
10	Glu Ala Lys Asp Val Phe Leu Gly Met Phe Leu Tyr Glu Tyr Ala Arg		
	325	330	335
	Arg His Pro Asp Tyr Ser Val Val Leu Leu Leu Arg Leu Ala Lys Thr		
	340	345	350
15	Tyr Glu Thr Thr Leu Glu Lys Cys Cys Ala Ala Ala Asp Pro His Glu		
	355	360	365
	Cys Tyr Ala Lys Val Phe Asp Glu Phe Lys Pro Leu Val Glu Glu Pro		
	370	375	380
	Gln Asn Leu Ile Lys Gln Asn Cys Glu Leu Phe Lys Gln Leu Gly Glu		
	385	390	395
20	Tyr Lys Phe Gln Asn Ala Leu Leu Val Arg Tyr Thr Lys Lys Val Pro		
	405	410	415
	Gln Val Ser Thr Pro Thr Leu Val Glu Val Ser Arg Asn Leu Gly Lys		
	420	425	430
	Val Gly Ser Lys Cys Cys Lys His Pro Glu Ala Lys Arg Met Pro Cys		
25	435	440	445
	Ala Glu Asp Tyr Leu Ser Val Val Leu Asn Gln Leu Cys Val Leu His		
	450	455	460
	Glu Lys Thr Pro Val Ser Asp Arg Val Thr Lys Cys Cys Thr Glu Ser		
	465	470	475
30	Leu Val Asn Arg Arg Pro Cys Phe Ser Ala Leu Glu Val Asp Glu Thr		
	485	490	495
	Tyr Val Pro Lys Glu Phe Asn Ala Glu Thr Phe Thr Phe His Ala Asp		
	500	505	510
35	Ile Cys Thr Leu Ser Glu Lys Glu Arg Gln Ile Lys Lys Gln Thr Ala		
	515	520	525
	Leu Val Glu Leu Val Lys His Lys Pro Lys Ala Thr Lys Glu Gln Leu		
	530	535	540
	Lys Ala Val Met Asp Asp Phe Ala Ala Phe Val Glu Lys Cys Cys Lys		
	545	550	555
40	Ala Asp Asp Lys Glu Thr Cys Phe Ala Glu Glu Gly Lys Lys Leu Val		
	565	570	575
	Ala Ala Ser Gln Ala Ala Leu Gly Leu		
	580	585	

45 (2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1865 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

50 (vii) IMMEDIATE SOURCE:

- (B) CLONE: native coding sequence of mature HSA

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

	AGATGCACAC AAGAGTGAGG TTGCTCATCG GTTAAAGAT TTGGGAGAAG AAAATTCAA	60
	AGCCTTGGTG TTGATTGCCT TTGCTCAGTA TCTTCAGCAG TGTCCATTG AAGATCATGT	120
60	AAAATTAGTG AATGAAGTAA CTGAATTGCA AAAAACATGT GTAGCTGATG AGTCAGCTGA	180
	AAATTGTGAC AAATCACTTC ATACCCTTT TGGAGACAAA TTATGCACAG TTGCAACTCT	240
	TCGTGAAACC TATGGTGAAA TGGCTGACTG CTGTGCAAAA CAAGAACCTG AGAGAAATGA	300
	ATGCTTCTTG CAACACAAAG ATGACAAACCC AAACCTCCC CGATTGGTGA GACCAGAGGT	360
	TGATGTGATG TGCAGTGCCTT TTCATGACAA TGAAGAGACA TTTTGAAAAA AATACTTATA	420
65	TGAAATTGCC AGAAGACATC CTTACTTTA TGCCCCGGAA CTCCCTTTCT TTGCTAAAAG	480
	GTATAAAAGCT GCTTTACAG AATGTTGCCA AGCTGCTGAT AAAGCTGCCT GCCTGTTGCC	540
	AAAGCTCGAT GAACTTCGGG ATGAAGGGAA GGCTTCGTCT GCCAACACAGA GACTCAAATG	600

	TGCCAGTCTC CAAAAATTG GAGAAAGAGC TTTCAAAGCA TGGGCAGTGG CTCGCCTGAG	660
	CCAGAGATT CCCAAAGCTG AGTTTGAGA AGTTTCCAAG TTAGTGACAG ATCTTACCAA	720
	AGTCCACACG GAATGCTGCC ATGGAGATCT GCTTGAATGT GCTGATGACA GGGCGGACT	780
5	TGCCAAGTAT ATCTGTGAAA ATCAGGATTC GATCTCCAGT AAACTGAAGG AATGCTGTGA	840
	AAAACCTCTG TTGGAAAAAT CCCACTGCAT TGCCGAAGTG GAAAATGATG AGATGCCTGC	900
	TGACTTGCTC TCATAGCTG CTGATTTGT TGAAAGTAAG GATGTTGCA AAAACTATGC	960
	TGAGGCAAAG GATGCTTCC TGGGCATGTT TTGATGAA TATGCAAGAA GGCATCCTGA	1020
	TTACTCTGTC GTGCTGCTG TGAGACTGCA AGAACATAT GAAACCACTC TAGAGAAGTG	1080
10	CTGTGCCGCT GCAGATCCTC ATGAATGCTA TGCCAAAGTG TTCGATGAAAT TAAACCTCT	1140
	TGTGGAAGAG CCTCAGAATT TAATCAAACA AAACTGTGAG CTTTTAAGC AGCTTGGAGA	1200
	GTACAAATTC CAGAATGCGC TATTAGTTCG TTACACCAAG AAAGTACCCC AAGTGTCAAC	1260
	TCCAACCTT GTAGAGGTCT CAAGAAACCT AGGAAAAGTG GGCAGCAAAT GTTGTAAACA	1320
	TCCGTGAAGCA AAAAGAATGC CCTGTGCAGA AGACTATCTA TCCGTGGTCC TGAACCAGTT	1380
15	ATGTGTGTTG CATGAGAAAA CGCCAGTAAG TGACAGAGTC ACAAATGCT GCACAGAGTC	1440
	CTTGGTGAAC AGGGGACCAT GCTTTTCAGC TCTGGAAGTC GATGAAACAT ACGTTCCCAA	1500
	AGAGTTAAAT GCTGAAACAT TCACCTTCCA TGCAAGATATA TGCACACTTT CTGAGAAGGA	1560
	GAGACAAATC AAGAAACAAA CTGCACTGTG TGAGCTTGTG AAACACAAAGC CCAAGGCAAC	1620
	AAAAGAGCAA CTGAAAGCTG TTATGGATGA TTTCGCAGCT TTTGTAGAGA AGTGCTGCAA	1680
20	GGCTGACGAT AAGGGAGACCT GCTTTGCCGA GGAGGGTAAA AAACTTGTTG CTGCAAGTCA	1740
	AGCTGCCTA GGCTTATAAC ATCTACATT AAAAGCATCT CAGCCTACCA TGAGAATAAG	1800
	AGAAAGAAAA TGAAGATCAA AAGCTTATTC ATCTGTTTC TTTTCGTTG GTGTAAAGGCC	1860
	AACAC	1865

(2) INFORMATION FOR SEQ ID NO:13:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 352 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: protein
 (vii) IMMEDIATE SOURCE:
 (B) CLONE: native proBPN' amino acid sequence

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Ala	Gly	Lys	Ser	Asn	Gly	Glu	Lys	Tyr	Ile	Val	Gly	Phe	Lys	Gln	
1		5					10						15		
Thr	Met	Ser	Thr	Met	Ser	Ala	Ala	Lys	Lys	Lys	Asp	Val	Ile	Ser	Glu
						20		25					30		
Lys	Gly	Gly	Lys	Val	Gln	Lys	Gln	Phe	Lys	Tyr	Val	Asp	Ala	Ala	Ser
						35		40					45		
Ala	Thr	Leu	Asn	Glu	Lys	Ala	Val	Lys	Glu	Leu	Lys	Lys	Asp	Pro	Ser
						50		55					60		
Val	Ala	Tyr	Val	Glu	Glu	Asp	His	Val	Ala	His	Ala	Tyr	Ala	Gln	Ser
						65		70				75			80
Val	Pro	Tyr	Gly	Val	Ser	Gln	Ile	Lys	Ala	Pro	Ala	Leu	His	Ser	Gln
						85		90				95			
Gly	Tyr	Thr	Gly	Ser	Asn	Val	Lys	Val	Ala	Val	Ile	Asp	Ser	Gly	Ile
						100		105				110			
Asp	Ser	Ser	His	Pro	Asp	Leu	Lys	Val	Ala	Gly	Gly	Ala	Ser	Met	Val
						115		120				125			
Pro	Ser	Glu	Thr	Asn	Pro	Phe	Gln	Asp	Asn	Asn	Ser	His	Gly	Thr	His
						130		135				140			
55	Val	Ala	Gly	Thr	Val	Ala	Ala	Leu	Asn	Asn	Ile	Gly	Val	Leu	Gly
						145		150				155			160
Val	Ala	Pro	Ser	Ala	Ser	Leu	Tyr	Ala	Val	Lys	Val	Leu	Gly	Ala	Asp
						165		170				175			
Gly	Ser	Gly	Gln	Tyr	Ser	Trp	Ile	Ile	Asn	Gly	Ile	Glu	Trp	Ala	Ile
						180		185				190			
Ala	Asn	Asn	Met	Asp	Val	Ile	Asn	Met	Ser	Leu	Gly	Gly	Pro	Ser	Gly
						195		200				205			
Ser	Ala	Ala	Leu	Lys	Ala	Ala	Val	Asp	Lys	Ala	Val	Ala	Ser	Gly	Val
						210		215				220			
65	Val	Val	Val	Ala	Ala	Gly	Asn	Glu	Gly	Thr	Ser	Gly	Ser	Ser	Ser
						225		230				235			240
Thr	Val	Gly	Tyr	Pro	Gly	Lys	Tyr	Pro	Ser	Val	Ile	Ala	Val	Gly	Ala

	245	250	255
	Val Asp Ser Ser Asn Gln Arg Ala Ser	Phe Ser Ser Val Gly Pro Glu	
	260	265	270
5	Leu Asp Val Met Ala Pro Gly Val Ser	Ile Gln Ser Thr Leu Pro Gly	
	275	280	285
	Asn Lys Tyr Gly Ala Tyr Asn Gly Thr Ser Met Ala Ser Pro His Val		
	290	295	300
10	Ala Gly Ala Ala Ala Leu Ile Leu Ser Lys His Pro Asn Trp Thr Asn		
	305	310	315
	Thr Gln Val Arg Ser Ser Leu Glu Asn Thr Thr Lys Leu Gly Asp		
	325	330	335
	Ser Phe Tyr Tyr Gly Lys Gly Leu Ile Asn Val Gln Ala Ala Ala Gln		
	340	345	350

15 (2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1056 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: native proBPN' coding sequence

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

	GCAGGGAAAT CAAACGGGGA AAAGAAATAT ATTGTCGGGT TTAAACAGAC AATGAGCAGC	60
	ATGAGCGCCG CTAAGAAGAA AGATGTCATT TCTGAAAAAG GCGGGAAAGT GCAAAAGCAA	120
30	TTCAATATG TAGACGCAGC TTCAGCTACA TTAAACGAAA AAGCTGTAAA AGAATTGAAA	180
	AAAGACCCGA GCGTCGCTTA CGTTGAAGAA GATCACGTAG CACATGCGTA CGCGCAGTCC	240
	GTGCCTTACG GCGTATCACA ATTAAAGCC CCTGCTCTGC ACTCTCAAGG CTACACTGGA	300
	TCAAATGTTA AAGTAGCGGT TATCGACAGC GGTATCGATT CTTCTCATCC TGATTTAAAG	360
35	GTAGCAGGCG GAGCCAGCAT GGTTCTTCT GAAACAAATC CTTTCAAGA CAACAACCT	420
	CACGGAACTC ACGTTGCCG CACAGTGC GCTCTTAAATA ACTCAATCGG TGTATTAGGC	480
	GTTGCGCCAA GCGCATCACT TTACGCTGTA AAAGTTCTCG GTGCTGACGG TTCCGGCCAA	540
	TACAGCTGGA TCATTAACGG AATCGAGTGG GCGATCGCAA ACAATATGGA CGTTATTAAAC	600
40	ATGAGCCTCG GCGGACCTTC TGGTTCTGCT GCTTTAAAAG CGGCAGTTGA TAAAGCCGTT	660
	GCATCCGGCG TCGTAGTCGT TCGGGCAGCC GGTAAACGAAG GCACTTCCGG CAGCTCAAGC	720
	ACAGTGGGCT ACCCTGGTAA ATACCCTTCT GTCATTCGAG TAGGCGCTGT TGACAGCAGC	780
	AACCAAAGAG CATCTTCTC AAGCGTAGGA CCTGAGCTTG ATGTCATGGC ACCTGGCGTA	840
45	TCTATCCAAA GCACGCTTCC TGGAAACAAA TACGGGGCGT ACAACGGTAC GTCAATGGCA	900
	TCTCCGCAAG TTGCCGGAGC GGCTGCTTTG ATTCTTTCTA AGCACCCGAA CTGGACAAAC	960
	ACTCAAGTCC GCAGCAGTTT AGAAAACACC ACTACAAAAC TTGGTGATTTC TTTCTACTAT	1020
	GGAAAAGGGC TGATCAACGT ACAGGGCGCA GCTCAG	1056

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 77 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: subtilisin BPN' pro-peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

60	Ala Gly Lys Ser Asn Gly Glu Lys Lys Tyr Ile Val Gly Phe Lys Gln			
	1	5	10	15
	Thr Met Ser Thr Met Ser Ala Ala Lys Lys Lys Asp Val Ile Ser Glu			
	20	25	30	
65	Lys Gly Lys Val Gln Lys Gln Phe Lys Tyr Val Asp Ala Ala Ser			
	35	40	45	
	Ala Thr Leu Asn Glu Lys Ala Val Lys Glu Leu Lys Lys Asp Pro Ser			
	50	55	60	

Val Ala Tyr Val Glu Glu Asp His Val Ala His Ala Tyr
 65 70 75

5 (2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 275 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vii) IMMEDIATE SOURCE:

- (B) CLONE: native mature BPN' amino acid sequence

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Ala Gln Ser Val Pro Tyr Gly Val Ser Gln Ile Lys Ala Pro Ala Leu
 1 5 10 15
 20 His Ser Gln Gly Tyr Thr Gly Ser Asn Val Lys Val Ala Val Ile Asp
 20 25 30
 Ser Gly Ile Asp Ser Ser His Pro Asp Leu Lys Val Ala Gly Gly Ala
 35 40 45
 Ser Met Val Pro Ser Glu Thr Asn Pro Phe Gln Asp Asn Asn Ser His
 50 55 60
 Gly Thr His Val Ala Gly Thr Val Ala Ala Leu Asn Asn Ser Ile Gly
 65 70 75 80
 Val Leu Gly Val Ala Pro Ser Ala Ser Leu Tyr Ala Val Lys Val Leu
 85 90 95
 30 Gly Ala Asp Gly Ser Gly Gln Tyr Ser Trp Ile Ile Asn Gly Ile Glu
 100 105 110
 Trp Ala Ile Ala Asn Asn Met Asp Val Ile Asn Met Ser Leu Gly Gly
 115 120 125
 Pro Ser Gly Ser Ala Ala Leu Lys Ala Ala Val Asp Lys Ala Val Ala
 130 135 140
 35 Ser Gly Val Val Val Val Ala Ala Ala Gly Asn Glu Gly Thr Ser Gly
 145 150 155 160
 Ser Ser Ser Thr Val Gly Tyr Pro Gly Lys Tyr Pro Ser Val Ile Ala
 165 170 175
 40 Val Gly Ala Val Asp Ser Ser Asn Gln Arg Ala Ser Phe Ser Ser Val
 180 185 190
 Gly Pro Glu Leu Asp Val Met Ala Pro Gly Val Ser Ile Gln Ser Thr
 195 200 205
 Leu Pro Gly Asn Lys Tyr Gly Ala Tyr Asn Gly Thr Ser Met Ala Ser
 210 215 220
 45 Pro His Val Ala Gly Ala Ala Ala Leu Ile Leu Ser Lys His Pro Asn
 225 230 235 240
 Trp Thr Asn Thr Gln Val Arg Ser Ser Leu Glu Asn Thr Thr Lys
 245 250 255
 50 Leu Gly Asp Ser Phe Tyr Tyr Gly Lys Gly Leu Ile Asn Val Gln Ala
 260 265 270
 Ala Ala Gln
 275

55 (2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 275 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vii) IMMEDIATE SOURCE:

- (B) CLONE: amino acid sequence of mature BPN' variant

65

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

5 Ala Gln Ser Val Pro Tyr Gly Val Ser Gln Ile Lys Ala Pro Ala Leu
 1 5 10 15
 His Ser Gln Gly Tyr Thr Gly Ser Asn Val Lys Val Ala Val Ile Asp
 20 25 30
 Ser Gly Ile Asp Ser Ser His Pro Asp Leu Lys Val Ala Gly Gly Ala
 35 40 45
 10 Ser Met Val Pro Ser Glu Thr Asn Pro Phe Gln Asp Thr Asn Ser His
 50 55 60
 Gly Thr His Val Ala Gly Thr Val Ala Ala Leu Thr Asn Ser Ile Gly
 65 70 75 80
 Val Leu Gly Val Ala Pro Ser Ala Ser Leu Tyr Ala Val Lys Val Leu
 15 85 90 95
 Gly Ala Asp Gly Ser Gly Gln Tyr Ser Trp Ile Ile Asn Gly Ile Glu
 100 105 110
 Trp Ala Ile Ala Asn Asn Met Asp Val Ile Thr Met Ser Leu Gly Gly
 115 120 125
 20 Pro Ser Gly Ser Ala Ala Leu Lys Ala Ala Val Asp Lys Ala Val Ala
 130 135 140
 Ser Gly Val Val Val Val Ala Ala Ala Gly Asn Glu Gly Thr Ser Gly
 145 150 155 160
 Ser Ser Ser Thr Val Gly Tyr Pro Gly Lys Tyr Pro Ser Val Ile Ala
 165 170 175
 25 Val Gly Ala Val Asp Ser Ser Asn Gln Arg Ala Ser Phe Ser Ser Val
 180 185 190
 Gly Pro Glu Leu Asp Val Met Ala Pro Gly Val Ser Ile Gln Ser Thr
 195 200 205
 30 Leu Pro Gly Asn Lys Tyr Gly Ala Tyr Ser Gly Thr Ser Met Ala Ser
 210 215 220
 Pro His Val Ala Gly Ala Ala Ala Leu Ile Leu Ser Lys His Pro Thr
 225 230 235 240
 Trp Thr Asn Thr Gln Val Arg Ser Ser Leu Glu Asn Thr Thr Thr Lys
 245 250 255
 35 Leu Gly Asp Ser Phe Tyr Tyr Gly Lys Gly Leu Ile Asn Val Gln Ala
 260 265 270
 Ala Ala Gln
 275

40 (2) INFORMATION FOR SEQ ID NO:18:

45 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1260 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

50 (vii) IMMEDIATE SOURCE:

- (B) CLONE: codon-optimized 3D signal peptide-AAT DNA sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

55 ATGAAGAACACCTCCCTCTGCCTCCTGCTGCTCGTGGTCCTCTGCTCCTGACCTGC 60
 AACAGCGGCCAGGCCGAGGA CCCGCAGGGC GACGCCGCCAGAAGACCGAACCCAGCCAC 120
 CACGACCAGG ACCACCCGAC GTTCAACAAG ATCACCCCCGA ATTTGGCCGA ATTTCGCCTTC 180
 AGCCTGTACC GCCAGCTCGC GCACCAAGTCC AACTCCACCA ACATCTTCTT CAGCCGGTG 240
 AGCATCGCCA CGCCCTTCGC CATGCTGTCC CTGGGTACCA AGGCGGACAC CCACGACGAG 300
 ATCCTCGAAG GGCTGAACCT CAACCTGACG GAGATCCGG AGGCGCAGAT CCACGAGGGC 360
 60 TTCCAGGAGCTGCTCAGGAC GCTCAACCAG CCGGACTCTCCAGCTCCAGCT CACCACCGGC 420
 AACGGGCTCTTCCTGTCCGA GGGCCTCAAG CTCGTCGATA AGTTCTGGAGGACGTGAAG 480
 AAGCTCTACC ACTCCGAGGC GTTCACCGTC AACTTCGGGG ACACCGAGGA GGCAAGAAG 540
 CAGATCAACG ACTACGTCGA GAAGGGGACCCAGGGCAAGA TCCTGGACCT GGTCAAGGAA 600
 TTGGACAGGG ACACCGTCTT CGCGCTCGTC AACTACATCT TCTTCAAGGG CAAGTGGGAG 660
 65 CGCCCGTTCAGGTGAAGGA CACCGAGGGAG GAGGACTTCAACGTCGACCA GGTCAACCAC 720
 GTCAAGGTCC CGATGATGAA GAGGCTCGGC ATGTTCAACA TCCAGCACTG CAAGAAGCTC 780
 TCCAGCTGGG TGCTCCTCAT GAAGTACCTG GGGAACGCCA CCGCCATCTT CTTCCCTGCCG 840

5	GACGAGGGCA AGCTCCAGCA CCTGGAGAAC GAGCTGACGC ACGACATCAT CACGAAGTTC	900
	CTGGAGAACG AGGACAGGCG CTCCGCTAGC CTCCACCTCC CGAAGCTGAG CATCACCGGC	960
	ACGTACGACC TGAAGAGCGT GCTGGCCAG CTGGGCATCA CGAAGGTCTT CAGCAACGGC	1020
	GCGGACCTCT CCGGCGTGCAG GGAGGAGGCC CCCCTGAAGC TCTCCAAGGC CGTGCACAAG	1080
	GCGGTGCTCA CGATCGACGA GAAGGGACG GAAGCTGCCG GGGCCATGTT CCTGGAGGCC	1140
	ATCCCCATGT CCATCCCGCC CGAGGTCAAG TTCAACAAAGC CCTTCGTCTT CCTGATGATC	1200
	GAGCAGAACCA CGAAGAGGCC CCTCTTCATG GGGAAAGGTGCG TCAACCCCCAC GCAGAAGTGA	1260

(2) INFORMATION FOR SEQ ID NO:19:

10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1382 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15

(vii) IMMEDIATE SOURCE:

- (B) CLONE: codon-optimized 3D signal peptide-ATIII DNA sequen

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

25	ATGAAGAACAA CCTCCTCCCT CTGCCTCCTG CTGCTCGTGG TCCTCTGCTC CCTGACCTGC	60
	AACAGCGGCC AGGCCACGG AAGCCCTGTG GACATCTGCA CAGCCAAGCC GCGGGACATT	120
	CCCATGAATC CCATGTGCAT TTACCGCTCC CGGGAGAAGA AGGCAACTGTA GGATGAGGGC	180
	TCAGAACAGA AGATCCCGGA GGCCACCAAC CGGCGTGTCT GGGAACTGTC CAAGGCCAAT	240
	TCCCCGCTTTG CTACCACTTT CTATCAGCAC CTGGCAGATT CCAAGAAATGA CAATGATAAC	300
	ATTTCCCTGT CACCCCTGAG TATCTCCACG GCTTTGCTA TGACCAAGCT GGGTGCCTGT	360
	AATGACACCC TCCAGCAACT GATGGAGGTA TTTAAGTTTG ACACCATATC TGAGAAAACA	420
	TCTGATCAGA TCCACCTTCTT CTTTGCCAAA CTGAAC TGCC GACTCTATCG AAAAGCCAAC	480
30	AAATCCTCA AGTTAGTATC AGCCAATCGC CTTTTGGAG ACAAACTCCCT TACCTTCAT	540
	GAGACCTACC AGGACATCAG TGAGTTGGTA TATGGAGCCA AGCTCCAGCC CCTGGACTTC	600
	AAGGAAAATG CAGAGCAATC CAGAGCGGCC ATCAACAAAT GGGTGTCCAA TAAGACCGAA	660
	GGCCGAATCA CCGATGTCAT TCCCTGGAA GCCATCAATG AGCTCACTGT TCTGGTGCCTG	720
	GTAAACACCA TTTACTTCAA GGGCCTGTGG AAGTCAAAGT TCAGCCTGAA GAACACAAGG	780
35	AAGGAACCTGT TCTACAAGGC TGATGGAGAG TCGTGTTCAG CATCTATGAT GTACCAGGAA	840
	GGCAAGTCC GTTATCGCG CGTGGCTGAA GGCACCCAGG TGCTTGAGTT GCCCCTCAA	900
	GGTGATGACA TCACCATGGT CCTCATCTTG CCCAAGCCTG AGAAGAGCCT GGCCAAGGTG	960
	GAGAAGGAAC TCACCCAGA GGTGCTGCAG GAGTGGCTGG ATGAATTGGA GGAGATGATG	1020
40	CTGGTGGTTC ACATGCCCG CTTCCGCATT GAGGACGGCT TCAGTTGAA GGAGCAGCTG	1080
	CAAGACATGG GCCTGTCGA TCTGTTTCAGC CCTGAAAAGT CCAAACCTCCC AGGTATTGTT	1140
	GCAGAAGGCC GAGATGACCT CTATGTCATCA GATGCATTCC ATAAGGCATT TCTTGAGGT	1200
	AATGAAGAAG GCAGTGAAGC AGCTGCAAGT ACCGCTGTTG TGATTGCTGG CGGTTCGCTA	1260
	AACCCCAACA GGGTGAACCTT CAAGGCCAAC AGGCCCTTC TGGTTTTAT AAGAGAAGTT	1320
45	CCTCTGAACCA CTATTATCTT CATGGGCAGA GTAGCCAACC CTTGTGTTAA GTAACCTCGAG	1380
	CC	1382

(2) INFORMATION FOR SEQ ID NO:20:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1940 base pairs.
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

55

(vii) IMMEDIATE SOURCE:

- (B) CLONE: codon-optimized 3D signal peptide-HSA DNA sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

60	ATGAAGAACAA CCTCCTCCCT CTGCCTCCTG CTGCTCGTGG TCCTCTGCTC CCTGACCTGC	60
	AACAGCGGCC AGGCCAGATG CACACAAGAG TGAGGTTGCT CATCGTTTA AAGATTGAG	120
	AGAAGAAAAT TTCAAAGCCT TGGTGTGAT TCCCTTGCT CAGTATCTTC AGCAGTGTCC	180
	ATTTGAAGAT CATCAAAT TAGTGAAGA AGTAACCTGAA TTGCAAAAAA CATGTGTAGC	240
	TGATGAGTCA GCTGAAAATT GTGACAAATC ACTTCATACC CTTTTGGAG ACAAAATTATG	300
65	CACAGTTGCA ACTCTCGTG AAACTATGG TGAATGGCT GACTGCTGTG CAAAACAAGA	360
	ACCTGAGAGA AATGAATGCT TCTTGCAACA CAAAGATGAC AACCCAAACC TCCCCCGATT	420
	GGTGAGACCA GAGGTTGATG TGATGTGCAC TGCTTTCAT GACAATGAAG AGACATTTT	480

	GAAAAAAATAC TTATATGAAA TTGCCAGAAG ACATCCTTAC TTTTATGCCCGGAACTCCT	540
	TTTCTTTGCT AAAAGGTATA AAGCTGCTTT TACAGAATGT TGCCAAGCTG CTGATAAAAGC	600
	TGCCTGCCCTG TTGCCAAAGC TCGATGAACT TCGGGATGAA GGGAAAGGCTT CGTCTGCCAA	660
5	ACAGAGACTC AAATGTGCCA GTCTCCAAA ATTGGAGAA AGAGCTTCA AAGCATGGC	720
	AGTGGCTCGC CTGAGCCAGA GATTTCCCAA AGCTGAGTT GCAGAAGTTT CCAAGTTAGT	780
	GACAGATCTT ACCAAAGTCC ACACGGAATG CTGCCATGGA GATCTGCTTG AATGTGCTGA	840
	TGACAGGGCG GACCTTGCA AGTATATCTG TGAAAATCAG GATTGATCT CCAGTAAACT	900
	GAAGGAATGC TGTGAAAAC CTCTGTTGGA AAAATCCCAC TGCAATTGCG AAGTGGAAA	960
10	TGATGAGATG CCTGCTGACT TGCCTTCATT AGCTGCTGAT TTTGTTGAAA GTAAGGATGT	1020
	TTGCAAAAAC TATGCTGAGG CAAAGGATGT CTTCTGGGC ATGTTTTGT ATGAAATATG	1080
	AAGAAGGCAT CCTGATTACT CTGTCGTGCT GCTGCTGAGA CTTGCCAAGA CATATGAAAC	1140
	CACTCTAGAG AAGTGTGTG CCGCTGCAGA TCCTCATGAA TGCTATGCCA AAGTGTTCGA	1200
	TGAATTAAA CCTCTGTGG AAGAGCCTCA GAATTTAATC AAACAAAATC GTGAGCTTT	1260
15	TAAGCAGCTT GGAGAGTACA ATTCCAGAA TGCGCTATTG TTGCGTTACA CCAAGAAAGT	1320
	ACCCCAAGTG TCAACTCCAA CTCTTGTAGA GGTCTCAAGA AACCTAGGAA AAGTGGGCAG	1380
	CAAATGTTGT AAACATCTG AAGCAAAAG AATGCCCTGT GCAGAAGACT ATCTATCCGT	1440
	GGTCTGAAC CAGTATGTG TGTGCTGATG GAAACAGCCA GTAAGTGACA GAGTCACAAA	1500
	ATGCTGCACA GAGTCTGTG TGAACAGGCG ACCATGCTTT TCAGCTCTGG AAGTCGATGA	1560
20	AACATACGTT CCCAAAGAGT TTAATGCTGA AACATTCAACC TTCCATGCGAT ATATATGCAC	1620
	ACTTTCTGAG AAGGAGAGAC AAATCAAGAA ACAAACTGCA CTTGTTGAGC TTGTGAAACA	1680
	CAAGCCCAAG GCAACAAAG AGCAACTGAA AGCTGTTATG GATGATTTCG CAGTTTTGT	1740
	AGAGAAGTGC TGCAAGGCTG ACGATAAGGA GACCTGCTT GCCGAGGAGG GTAAAAAAACT	1800
	TGTTGCTGCA AGTCAAGCTG CCTTAGGCTT ATAACATCTA CATTAAAAG CATCTCAGCC	1860
25	TACCATGAGA ATAAGAGAAA GAAAATGAAG ATCAAAAGCT TATTGATCTG TTTTCTTTT	1920
	CGTTGGTGTAAAGCCAAACAC	1940

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 1140 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

35 (B) CLONE: codon-optimized 3D signal peptide-BPN' DNA sequene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

40	ATGAAGAACACCTCCTCCCT CTGCTCGTGG TCCTCTGCTC CCTGACCTGC	60
	ACACAGCGGCC AGGCCGCTGG CAAGAGCAAC GGGGAGAAGA AGTACATCGT CGGCTTCAG	120
	CAGACCATGA GCACCATGAG CGCCGCCAG AAGAAGGACG TCATCAGCGA GAAGGGCGGC	180
	AAGGTACAGA AGCAGTTCAA GTACGTGGAC CCCGCCAGCG CCACCCCTCAA CGAGAAGGCC	240
45	GTCAAGGAGC TGAAGAAGGA CCCGAGCGTC GCCTACGTCG AGGAGGACCA CGTCGCCAAC	300
	GCATATGCAC AGAGCGTCCC GTACGGCTC AGCCAGATCA AGGCCCCGGC CCTCCACAGC	360
	CAGGGCTACA CCGGCAGCAA CGTCAAGGTC GCCGTACATCG ACAGCGGCAT CGACAGCAGC	420
	CACCCGGACC TCAAGGTCGC CGGCGGAGCT AGCATGGTCC CGAGCGAGAC CAACCCGTC	480
	CAGGACACCA ACAGCCATGG CACCCACGTC GCCGGCACCG TCGCCGCCCT CACCAACAGC	540
50	ATCGGCGTCC TCGGGCTCGC CCCGAGCGCC AGCCTCTACG CCGTCAAGGT ACTCGGCGCC	600
	GACGGCAGCG GCCAGTACAG CTGGATCATC AACGGCATTG AGTGGGCCAT CGCCAACAAC	660
	ATGGGACGTC TACCCATGAG CCTCGGGCGC CCGAGCGGCA GCGCCGCCCT CAAGGCGGCC	720
	GTCGACAAGG CCGTCGCCAG CGCGTCGCGT GTCGTCGCCCG CCGCCGGCAA CGAGGGCACC	780
55	AGCGGGCAGCA GCAGCACCGT CGGTACCCCG GGAAAGTACCG CGAGCGTCAT CGCCGTCGGC	840
	GCCGTGGACA GCAGCAACCA GCGCGCGAGC TTACGAGCG TCGGCCCGGA GCTGGACGTC	900
	ATGGCCCGGG GCGTCAGCAT CCAGAGCACC CTCCCGGGCA ACAAGTACGG CGCCTACAGC	960
	GGCACCAAGCA TGGCAGGCC GCACGTCGCC GGCGCCGCTG CACTCATCCT CAGCAAGCAC	1020
	CCGACCTGGA CCAACACCCA GGTCCGCAGC AGCCTGGAGA ACACCACCAAC CAAGCTCGGC	1080
60	GACAGCTTCT ACTACGGCAA GGGCCTCATC AACGTCCAGG CCGCCGCCA GTGACTCGAG	1140

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

65 (A) LENGTH: 13 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:
 (B) CLONE: N-terminus of mature AAT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

5 Glu Asp Pro Gln Gly Asp Ala Ala Gln Lys Thr Asp Thr
 1 5 10

(2) INFORMATION FOR SEQ ID NO:23:

10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

20 GCTTGACCTG TAACTCGGGC CAGGCGAGCT

30

(2) INFORMATION FOR SEQ ID NO:24:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CGCCTAGCCC GAGTTACAGG TCAAGCAGCT

30

(2) INFORMATION FOR SEQ ID NO:25:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 37 base pairs
 (B) TYPE: nucleic acid
 40 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

45 AGCTCCATGG CCGTGGCTCG AGTCTAGACG CGTCCCC

37

(2) INFORMATION FOR SEQ ID NO:26:

50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

60 GGGGACGCGT CTAGACTCGA GCCACGGCCA TGG

33

(2) INFORMATION FOR SEQ ID NO:27:

65 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 35 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

5 GCATGCAGGT GCTGAACACC ATGGTGAACA AACAC 35

(2) INFORMATION FOR SEQ ID NO:28:

10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TTCTTGTCCC TTTCGGTCCT CATCGTCCTC CT 32

(2) INFORMATION FOR SEQ ID NO:29:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

30 TGGCCTCTCC TCCAACTTGA CAGCCGGGAG CT 32

(2) INFORMATION FOR SEQ ID NO:30:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TTCACCATGG TGTTCAGCAC CTGCATGCTG CA 32

(2) INFORMATION FOR SEQ ID NO:31:

45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

55 CGATGAGGAC CGAAAGGGAC AAGAAAGTGTT TG 32

(2) INFORMATION FOR SEQ ID NO:32:

60 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 35 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

65 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CCCGGCTGTC AAGTTGGAGG AGAGGCCAAG GAGGA

35

(2) INFORMATION FOR SEQ ID NO:33:

5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GAGGATCCCC AGGGAGATGC TGCCCAGAA

29

15

(2) INFORMATION FOR SEQ ID NO:34:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CGCGCTCGAG TTATTTTG GG GTGGGATTCA CCAC

34

IT IS CLAIMED:

1. A method of producing, in monocot plant cells, a mature heterologous protein selected from the group consisting of

5 (i) mature, glycosylated α_1 -antitrypsin (AAT) having the same N-terminal amino acid sequence as mature AAT produced in humans and a glycosylation pattern which increases serum halflife substantially over that of mature non-glycosylated AAT;

(ii) mature, glycosylated antithrombin III (ATIII) having the same N-terminal amino acid sequence as mature ATIII produced in humans;

10 (iii) mature human serum albumin (HSA) having the same N-terminal amino acid sequence as mature HSA produced in humans and having the folding pattern of native mature HSA as evidenced by its bilirubin-binding characteristics; and

(iv) mature, active subtilisin BPN' (BPN') having the same N-terminal amino acid sequence as BPN' produced in *Bacillus*;

15 the method comprising:

(a) obtaining monocot cells transformed with a chimeric gene having (i) a monocot transcriptional regulatory region, inducible by addition or removal of a small molecule, or during seed maturation, (ii) a first DNA sequence encoding the heterologous protein, and (iii) a second DNA sequence encoding a signal peptide, said first and second DNA sequences in translation-frame 20 and encoding a fusion protein, and wherein (i) the transcriptional regulatory region is operably linked to the second DNA sequence, and (ii) said signal peptide is effective to facilitate secretion of the mature heterologous protein from the transformed cells;

(b) cultivating the transformed cells under conditions effective to induce said transcriptional regulatory region, thereby promoting expression of the fusion protein and secretion of the mature 25 heterologous protein from the transformed cells; and

(c) isolating said mature heterologous protein produced by the transformed cells.

2. The method of claim 1, wherein said first DNA sequence encodes proBPN', said cultivating includes cultivating said transformed cells at a pH between 5-6 to promote expression 30 and secretion of proBPN' from the cells, and said isolating step includes incubating the proBPN' under conditions effective to allow the autoconversion of proBPN' to active mature BPN'.

3. The method of claim 1, wherein said first DNA sequence encodes mature BPN', and said method further includes:

35 transforming said cells with a second chimeric gene containing (i) a transcriptional

regulatory region inducible by addition or removal of a small molecule, or during seed maturation, (ii) a third DNA sequence encoding the pro-peptide moiety of BPN', and (iii) a fourth DNA sequence encoding a signal polypeptide, where said fourth DNA sequence is operably linked to said transcriptional regulatory region and said third DNA sequence, and where said signal polypeptide is in translation-frame with said pro-peptide moiety and is effective to facilitate secretion of expressed pro-peptide moiety from the transformed cells;

5 said cultivating step includes cultivating the transformed cells at a pH between 5-6 to promote expression and secretion of BPN' and the pro-peptide moiety from the cells;

10 and said isolating step includes incubating the BPN' and the pro-moiety under conditions effective to allow the conversion of BPN' to active mature BPN', and isolating the active mature BPN'.

4. The method of claim 1, wherein said signal peptide is the RAmy3D signal peptide having the amino acid sequence identified by SEQ ID NO:1.

15

5. The method of claim 1, wherein said second DNA sequence encodes the RAmy3D signal peptide (SEQ ID NO:1) and has the codon-optimized nucleotide sequence identified by SEQ ID NO:3.

20

6. The method of claim 1, wherein said signal peptide is the RAmy1A signal peptide having the amino acid sequence identified by SEQ ID NO:4.

25

7. The method of claim 1, wherein the second DNA sequence, the first DNA sequence, or both the second and the first DNA sequence, is codon-optimized for enhanced expression in said plant.

30

8. The method of claim 1, wherein said transcriptional regulatory region is a promoter derived from a rice or barley α -amylase gene selected from the group consisting of the RAmy1A, RAmy1B, RAmy2A, RAmy3A, RAmy3B, RAmy3C, RAmy3D, and RAmy3E, pM/C, gKAmy141, gKAmy155, Amy32b, and HV18 genes.

35

9. The method of claim 8, wherein the chimeric gene further comprises, between said transcriptional regulatory region and said second DNA coding sequence, the 5' untranslated region of an inducible monocot gene selected from the group consisting of RAmy1A, RAmy3B, RAmy3C, RAmy3D, HV18, and RAmy3E.

10. The method of claim 8, wherein said chimeric gene further comprises, downstream of the sequence encoding said fusion protein, the 3' untranslated region of an inducible monocot gene derived from a rice or barley α -amylase gene selected from the group consisting of the RAmy1A, 5 RAmy1B, RAmy2A, RAmy3A, RAmy3B, RAmy3C, RAmy3D, and RAmy3E, pM/C, gKAmy141, gKAmy155, Amy32b, and HV18 genes.

11. The method of claim 1, wherein said cultivating includes culturing the transformed plant cells in a sugar-free or sugar-depleted medium, the transcriptional regulatory region is derived from 10 the RAmy3E or RAmy3D gene, the 5' untranslated region is derived from the RAmy1A gene and has the sequence identified by SEQ ID NO:5, and the 3' untranslated region is derived from the RAmy1A gene.

12. The method of claim 1, wherein the transformed cells are aleurone cells of mature 15 seeds, the transcriptional regulatory region is upregulated by addition of a small molecule to promote seed germination, and said cultivating includes germinating said seeds, either in embryonated or de-embryonated form.

13. The method of claim 12, wherein the transcriptional regulatory region is a rice α -amyylase RAmy1A promoter or a barley HV18 promoter, and said small molecule is gibberellic acid. 20

14. A mature heterologous protein produced by the method of claim 1, wherein said protein is selected from the group consisting of:

(i) mature glycosylated α_1 -antitrypsin (AAT) having the same N-terminal amino acid 25 sequence as mature AAT produced in humans and having a glycosylation pattern which increases serum halflife substantially over that of non-glycosylated mature AAT;

(ii) mature glycosylated antithrombin III (ATIII) having the same N-terminal amino acid sequence as mature ATIII produced in humans; and

(iii) mature glycosylated subtilisin BPN' (BPN') having the same N-terminal amino acid 30 sequence as BPN' produced in *Bacillus*;

wherein said protein has a glycosylation pattern characteristic of proteins produced in said monocot plant.

15. The method of claim 1, wherein said monocot plant cells are transformed rice, barley, 35 corn, wheat, oat, rye, sorghum, or millet cells.

16. The method of claim 1, wherein said monocot plant cells are transformed rice or barley cells.

5 17. Plant cells capable of producing the mature heterologous protein according to the method of claim 1, wherein said cultivating includes culturing the transformed plant cells in a sugar-free or sugar-depleted medium, the transcriptional regulatory region is derived from the RAmy3E or RAmy3D gene, the 5' untranslated region is derived from the RAmy1A gene and has the sequence identified by SEQ ID NO:5, and the 3' untranslated region is derived from the RAmy1A gene.

10

18. Seeds capable of producing the mature heterologous protein according to the method of claim 1, wherein said transformed cells are aleurone cells, the transcriptional regulatory region is upregulated by addition of a small molecule to promote seed germination, and said cultivating includes germinating said seeds, either in embryonated or de-embryonated form.

15

3D Signal Peptide

non-codon optimized ATG AAG AAC ACC AGC AGC TTG TTG CTG CTC CTC GTG GTG CTC TGC AGC TTG ACC TGT AAC TCG TCG CAG GCG
 codon-optimized ATG AAG AAC ACC TCC CTC TGC CTC CTG CTC GTC CTC GTC TCC ACC TGC AAC AGC GGC CAG GCC
 amino acid sequence Met Lys Asn Thr Ser Ser Leu Cys Ser Leu Val Leu Val Val Cys Asn Ser Thr Cys Asn Ser Gly Gln Ala

Fig. 1

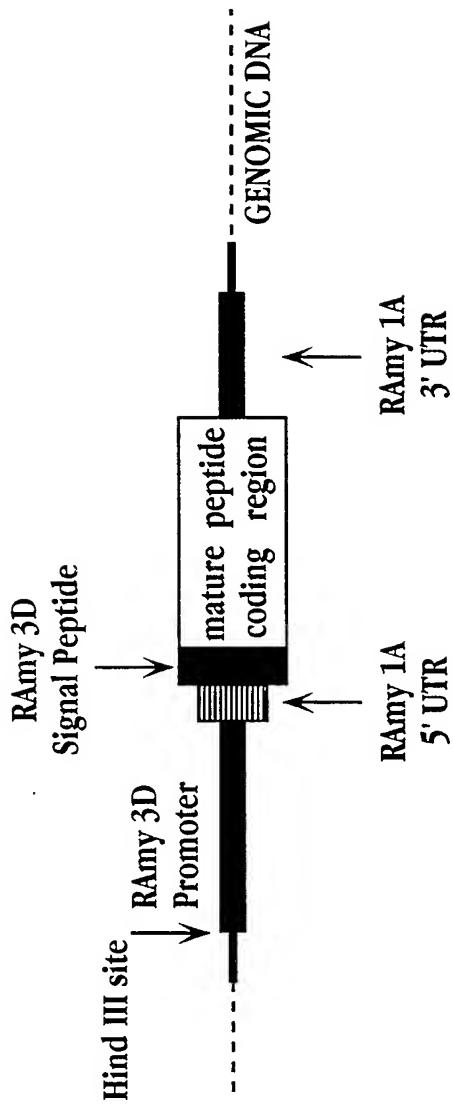
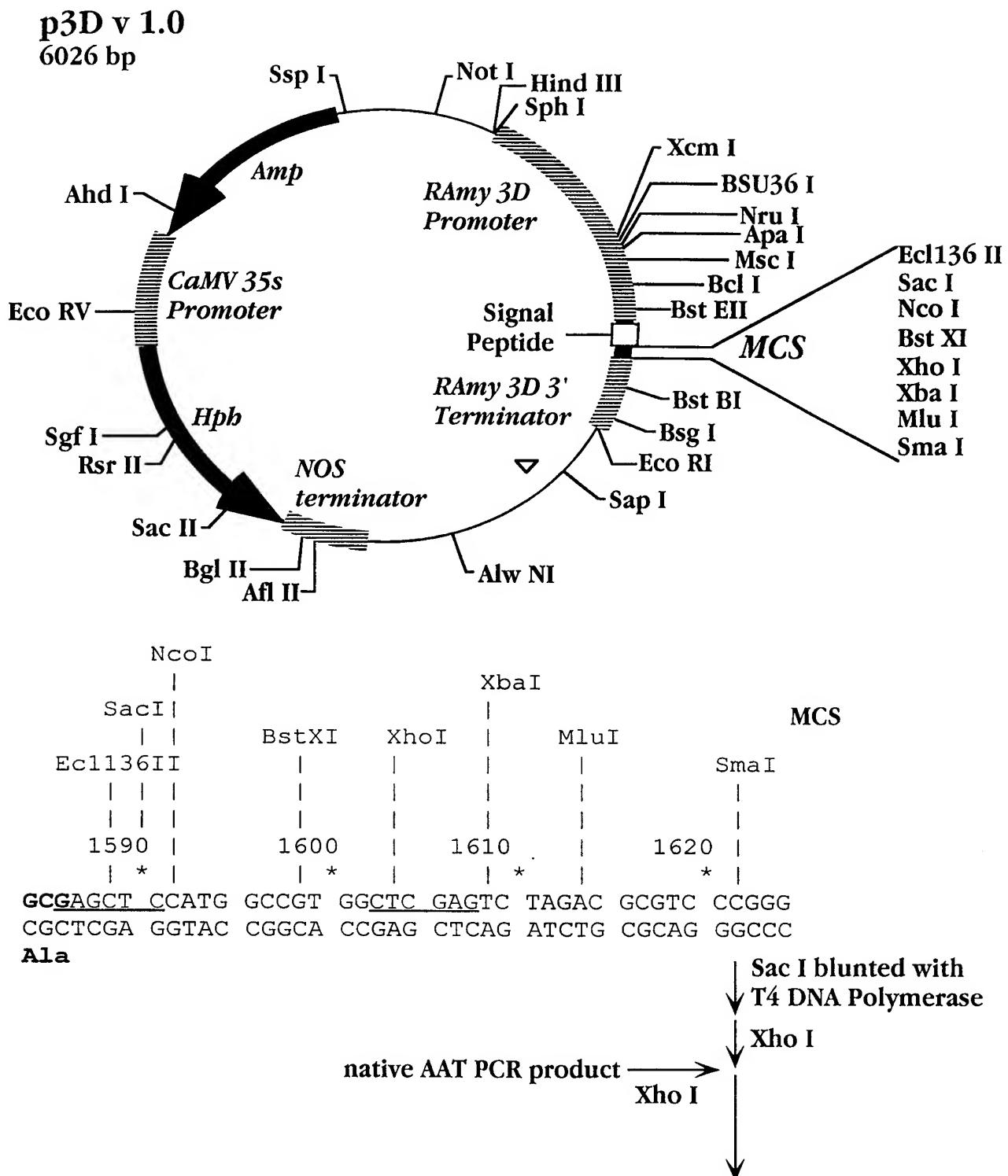
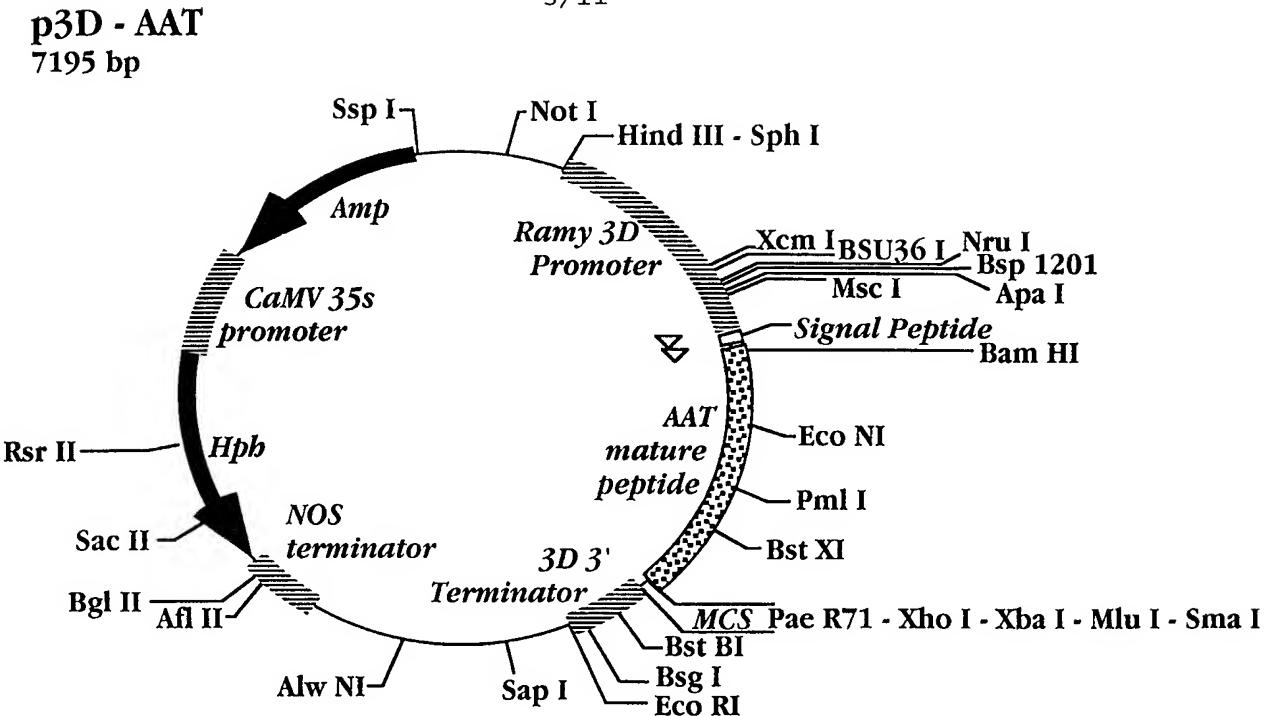


Fig. 2

2/11

**Fig. 3A**

3/11

**Fig. 3B**

4/11

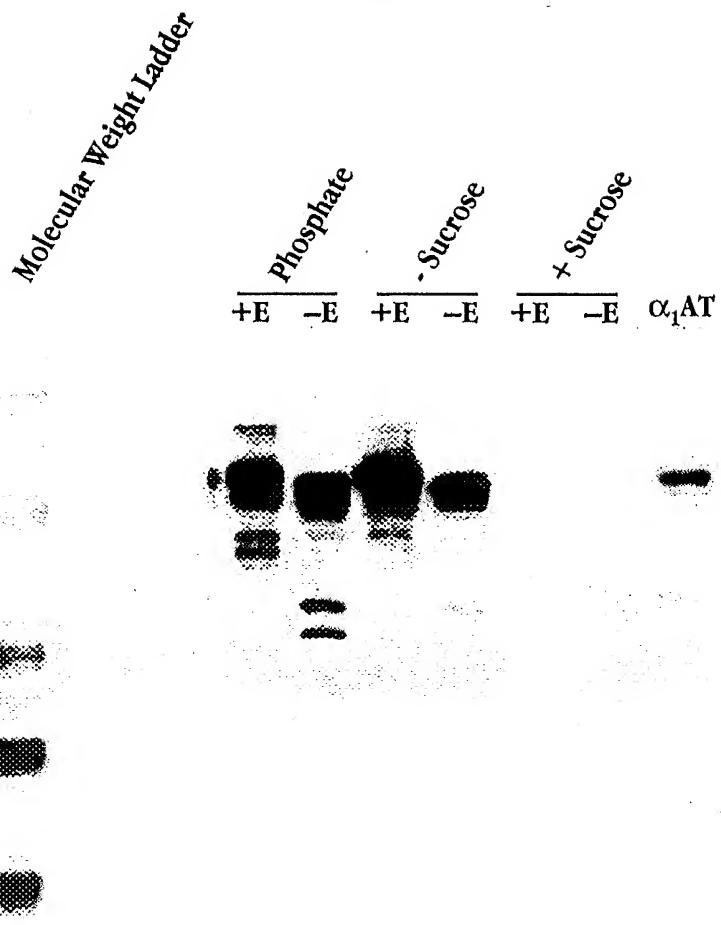


Fig. 4

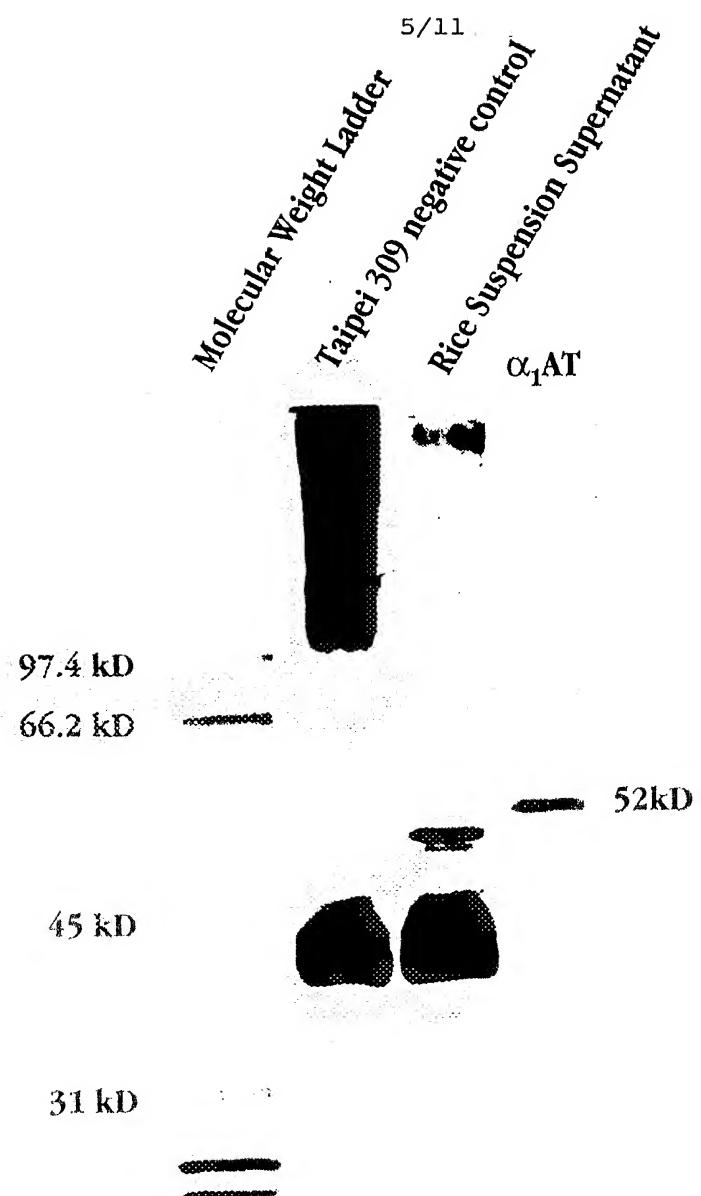


Fig. 5

6/11

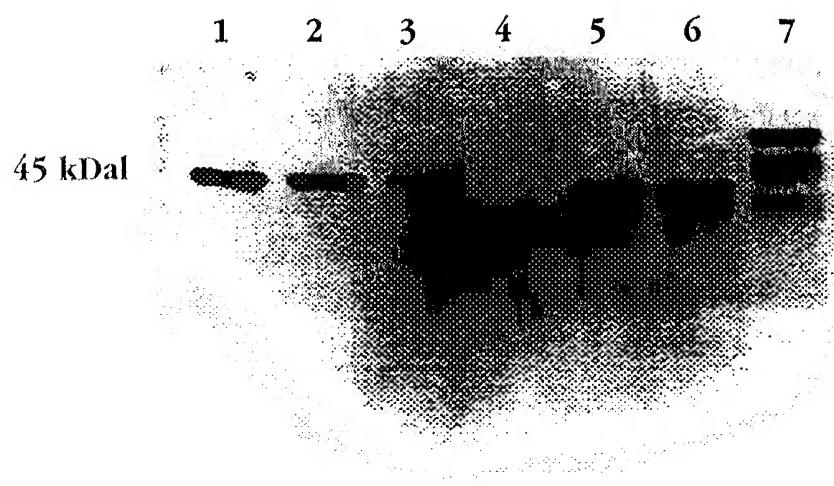


Fig. 6

7/11

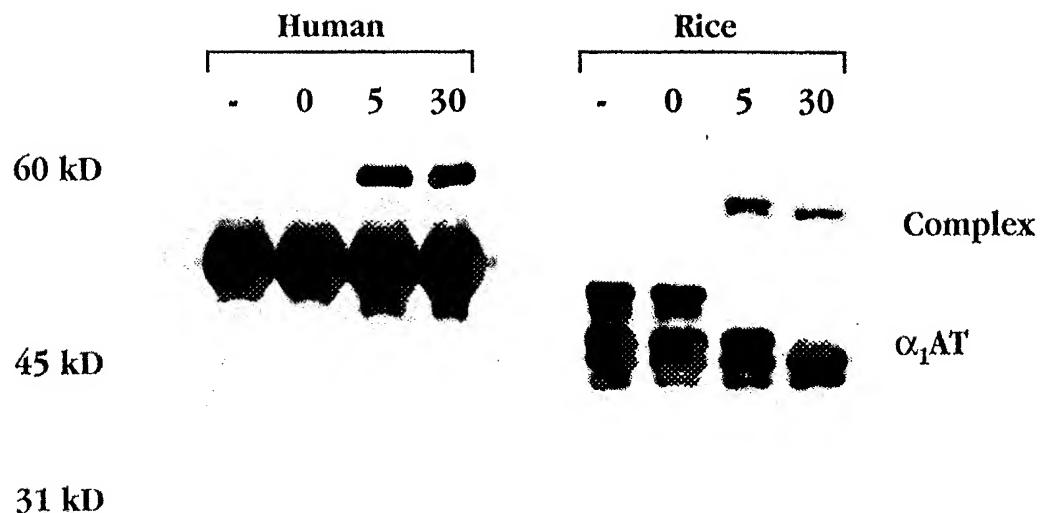


Fig. 7

N → E-D-P-Q-G-D-A-A-Q-K-T-D-T

Fig. 8

8/11

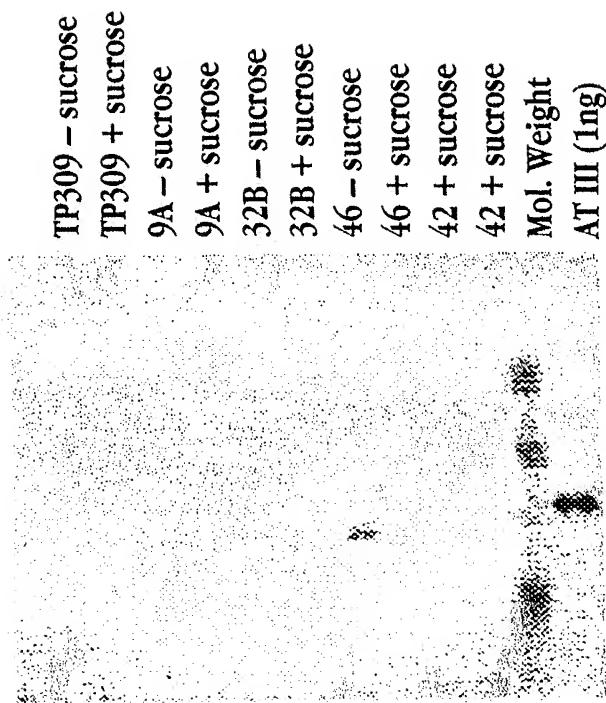


Fig. 9

9/11

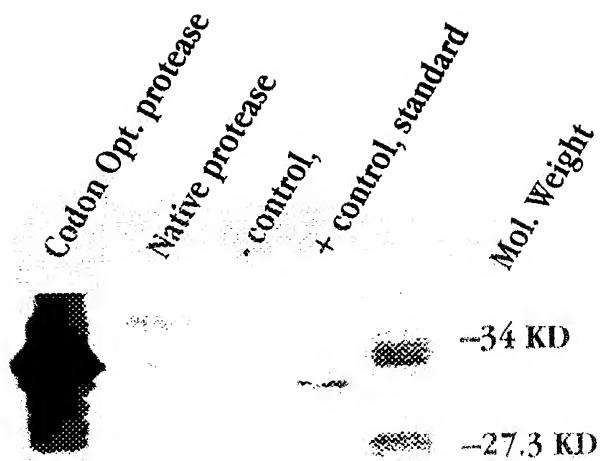
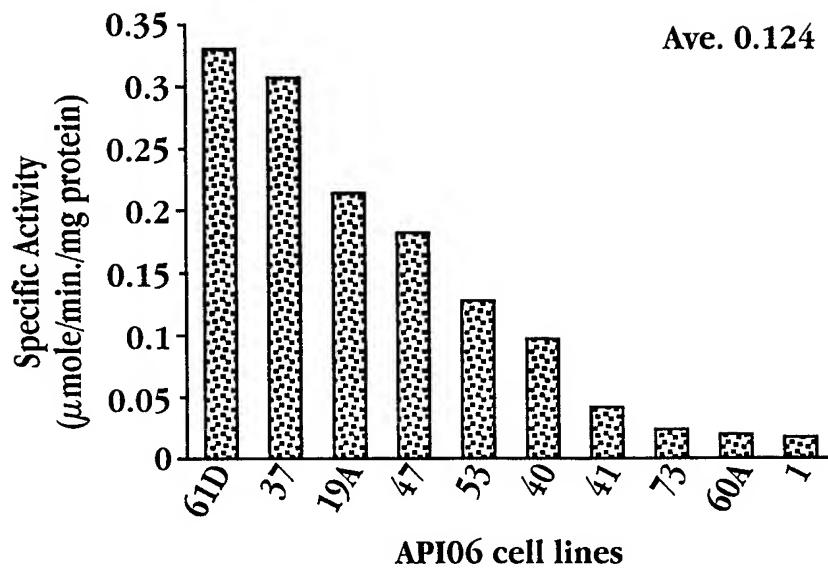
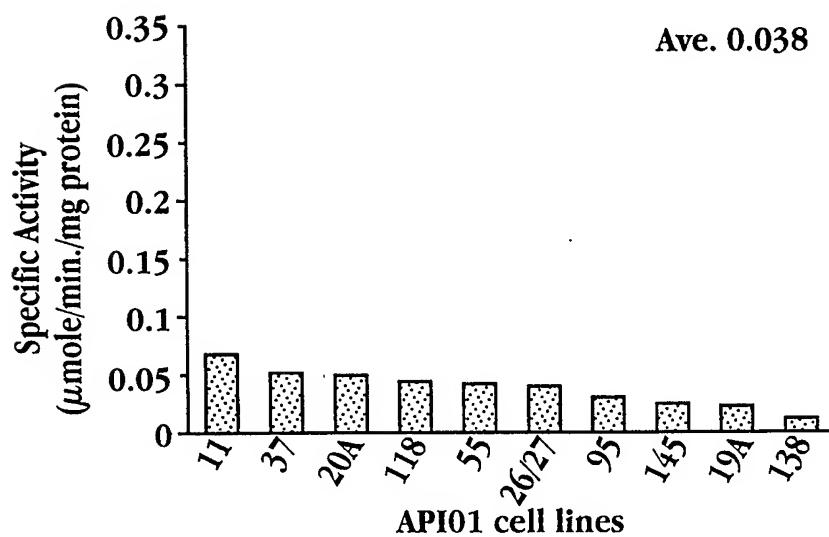


Fig. 10

10/11

**Fig. 11A****Fig. 11B**

11/11

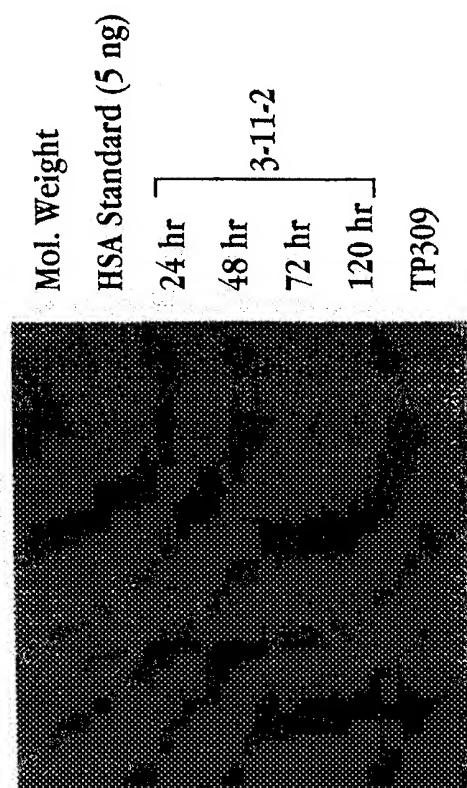


Fig. 12

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 98/03068

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/82 C12N15/57 C12N15/15 C12N15/14 C12P21/02

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category ^o	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 95 14099 A (RODRIGUEZ RAYMOND L ;UNIV CALIFORNIA (US)) 26 May 1995	1,4,6, 8-18
Y	see the whole document ---	1,5
Y	WO 92 01042 A (NOVONORDISK AS) 23 January 1992 see page 6, line 15 - line 19 ---	1
Y	JENSEN L G ET AL: "TRANSGENIC BARLEY EXPRESSING A PROTEIN-ENGINEERED, THERMOSTABLE (1,3-1,4)-BETA-FLUCANASE DURING GERMINATION" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 93, no. 8, April 1996, pages 3487-3491, XP002024710 see the whole document ---	5
	-/--	

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

^o Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

Date of mailing of the international search report

30 June 1998

14/07/1998

Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel: (+31-70) 340-2040, Tx. 31 651 epo nl
Fax: (+31-70) 340-3016

Authorized officer

Maddox, A

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 98/03068

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	TERASHIMA M ET AL: "Production of functional human alpha-1-antitrypsin by rice cell culture; expression and protein secretion in callus culture (conference abstract)" ABSTR.PAP.AM.CHEM.SOC.;(1997) 214 MEET., PT.1, AGFD018 CODEN: ACSRAL ISSN: 0065-7727 AMERICAN CHEMICAL SOCIETY, 214TH ACS NATIONAL MEETING, LAS VEGAS, NV, 7-11 SEPTEMBER, 1997., XP002069835 see abstract 018 ---	1,4, 8-10, 14-17
A	THOMAS, B. R. ET AL: "Gene regulation and protein secretion from plant cell cultures: the rice alpha - amylase system" ADVANCES IN PLANT BIOTECHNOLOGY, (1994) PP. 37-55. STUDIES IN PLANT SCIENCE 4. 85 REF. PUBLISHER: ELSEVIER SCIENCE. AMSTERDAM ISBN: 0-444-89939-1, XP002069833 see the whole document ---	1,11
A	CHAN M-T ET AL: "Novel gene expression system for plant cells based on induction of alpha-amylase promoter by carbohydrate starvation." JOURNAL OF BIOLOGICAL CHEMISTRY 269 (26). 1994. 17635-17641. ISSN: 0021-9258, XP002069821 see the whole document ---	1,11
A	US 5 460 952 A (YU SU-MAY ET AL) 24 October 1995 see the whole document ---	1,11
A	WO 90 01551 A (ROGERS JOHN C) 22 February 1990 see the whole document ---	1,12
A	WO 91 02066 A (MOGEN INT) 21 February 1991 see the whole document ---	1
A	EP 0 348 348 A (CIBA GEIGY AG) 27 December 1989 see examples 41-45 -----	1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 98/03068

Patent document cited in search report	Publication date	Patent family member(s)			Publication date
WO 9514099	A 26-05-1995	US 5693506	A	02-12-1997	
		AU 1289295	A	06-06-1995	
		CA 2176834	A	26-05-1995	
		EP 0788550	A	13-08-1997	
		JP 9509565	T	30-09-1997	
WO 9201042	A 23-01-1992	AU	8219291	A	04-02-1992
US 5460952	A 24-10-1995	JP	7143895	A	06-06-1995
		US	5712112	A	27-01-1998
WO 9001551	A 22-02-1990	AU	638409	B	01-07-1993
		AU	4037289	A	05-03-1990
		EP	0428572	A	29-05-1991
		JP	4500153	T	16-01-1992
		US	5677474	A	14-10-1997
WO 9102066	A 21-02-1991	NL	8901932	A	18-02-1991
		EP	0436003	A	10-07-1991
		JP	4502861	T	28-05-1992
		US	5650307	A	22-07-1997
		US	5716802	A	10-02-1998
		US	5763748	A	09-06-1998
EP 0348348	A 27-12-1989	AU	631551	B	03-12-1992
		AU	3656889	A	21-12-1989
		DK	302289	A	28-02-1990
		IL	90640	A	04-01-1998
		JP	2046238	A	15-02-1990